MODELING COVARIANCE STRUCTURE IN UNBALANCED LONGITUDINAL DATA

A Dissertation

by

MIN CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Statistics

MODELING COVARIANCE STRUCTURE IN UNBALANCED LONGITUDINAL DATA

A Dissertation

by

MIN CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Jianhua Huang
Committee Members,	Michael Sherman
	Suojin Wang
	Charles D. Phillips
Head of Department,	Simon J. Sheather

August 2008

Major Subject: Statistics

ABSTRACT

Modeling Covariance Structure in Unbalanced Longitudinal Data. (August 2008)

Min Chen, B.S., Xiamen University, P. R. China;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Jianhua Huang

Modeling covariance structure is important for efficient estimation in longitudinal data models. Modified Cholesky decomposition (Pourahmadi, 1999) is used as an unconstrained reparameterization of the covariance matrix. The resulting new parameters have transparent statistical interpretations and are easily modeled using covariates. However, this approach is not directly applicable when the longitudinal data are unbalanced, because a Cholesky factorization for observed data that is coherent across all subjects usually does not exist. We overcome this difficulty by treating the problem as a missing data problem and employing a generalized EM algorithm to compute the ML estimators. We study the covariance matrices in both fixed-effects models and mixed-effects models for unbalanced longitudinal data. We illustrate our method by reanalyzing Kenwards (1987) cattle data and conducting simulation studies.

This work is dedicated to my family.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Jianhua Huang, for his continuous support, guidance and encouragement not only on my research but also on my career development. His enthusiasm in statistics research always inspires me. I hope I can find a better word to express my most sincere appreciation.

I would like to thank the other members of my committee, Dr. Michael Sherman for being my master's advisor and giving me precious advice, Dr. Suojin Wang for his willingness to serve on my committee and giving me valuable comments, Dr. Charles Phillips for supporting my studies for the past four years.

I thank Dr. Michael Longnecker for encouraging me to pursue my Ph.D. degree. I also want to thank other professors in the department for giving me broad knowledge of statistics.

Finally, I would like to thank my family and all my friends for their devoted love and encouragement.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Modeling covariance matrices is important for efficient estimation and inference for means in regression with longitudinal data. Covariance matrices can be used in estimating means with a maximum likelihood approach. If covariance matrices are misspecified, inference for means is invalid using maximum likelihood method. Working covariance matrices can be used in estimation for means with generalized estimating equations (GEE). Inference for means would be valid even if working covariance matrices were misspecified in GEE. The estimation for means, however, would be inefficient, the confidence interval would be wider, the power of hypothesis tests would be lower. However, modeling the covariance matrices is challenging because a covariance matrix must be positive definite and it possibly has high dimensionality. The number of covariance matrix parameters grows quadratically with the size of the covariance matrix. The sample covariance matrix is generally positive definite and unbiased, but not stable due to its high dimensionality.

There have been several studies to overcome the problems of the positive definite constraint and the high dimensionality of such covariance matrices. Anderson (1973) proposed the class of linear covariance models (LCM)

$$\Sigma = \alpha_1 U_1 + \cdots + \alpha_q U_q,$$

in which $\alpha_i's$ are unknown parameters and $U_i's$ are known symmetric matrices. The

---

The format and style follow that of *Biometrics*.

$\alpha_i's$ are constrained such that Σ is positive-definite. Every covariance matrix can also be written as

$$\Sigma = (\sigma_{ij}) = \sum_{i=1}^{p}\sum_{j=1}^{p} \sigma_{ij} U_{ij},$$

in which U_{ij} is a $p \times p$ matrix with (i, j)-th entry being 1 and other entries being zero. To reduce the high-dimensionality of a covariance structure, certain entries of Σ or its inverse can be set to zero. However, the constraint on the coefficients makes parameter estimation difficult (Anderson, 1973; Szatrowski, 1980).

Leonard and Hsu (1992) and Chiu, Leonard, and Tsui (1996) introduced the log-linear covariance models which remove the constraint on $\alpha_i's$ in linear covariance models. For a general covariance matrix with the spectral decomposition $\Sigma = P\Lambda P'$, its matricial logarithm $\log \Sigma$ is defined by $\log \Sigma = P \log \Lambda P'$. Thus, a log-linear models for Σ is written as

$$\log \Sigma = \alpha_1 U_1 + \cdots + \alpha_q U_q,$$

in which $\alpha_i's$ are unknown parameters and $U_i's$ are known symmetric matrices. Unlike the linear covariance model, $\alpha_i's$ are unconstrained now. The drawback of log-linear covariance models is that the $\alpha_i's$ lack statistical interpretation since $\log \Sigma$ is a non-linear operation on Σ (Brown, Le, and Zidek, 1994).

The more successful approaches to model covariance structure are decomposing the matrix into its "dependence" and "variance " components.

The simplest decomposition of the covariance matrix is the variance-correlation decomposition given by

$$\Sigma = DRD,$$

in which R is the corresponding correlation matrix and $D = \text{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{\text{pp}}})$ with σ_{jj} standing for the j-th diagonal entry of Σ . Here, R and D can be interpreted

as "dependence" and "variance" components, respectively. Note that the correlation matrix R is positive-definite with constrained entries and logarithms of the diagonal entries of D are unconstrained when the diagonal entries are nonnegative.

Another decomposition of the covariance matrix is the spectral decomposition, which is also essential to the principal component analysis (Hotelling, 1933), factor analysis, and other techniques in multivariate statistics (Anderson, 2003). It's given by

$$\Sigma = P\Lambda P'$$

in which $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_\text{p})$ with λ_j being the j-th eigenvalue of Σ and $P's$ are orthogonal matrices. Here, P and Λ can be interpreted as "dependence" and "variance" components, respectively.

In both the variance-covariance decomposition and the spectral decomposition, there are constraints on the "dependence" components. Thus, it is inconvenient to use covariates to model them.

Pourahmadi (1999, 2000) provided unconstrained reparameterization of the covariance matrix in longitudinal studies. The reparameterization is based on a modified Cholesky decomposition of the covariance matrix Σ. The new parameters have transparent statistical interpretations in terms of the regression coefficients and logarithms of prediction error variances when regressing a response on its predecessors. In addition, the new parameters are easily modeled using covariates. With unconstrained reparameterization of the covariance matrix and modeling the new parameters using covariates, we can remove the difficulties of high dimensionality and the positive definite constraint in modeling a covariance matrix.

Following Pourahmadi's work, several studies have been done for modeling covariance matrices in longitudinal data with the modified Cholesky decomposition.

Pan and MacKenzie (2006) proposed methods that include baseline covariates in modeling covariance matrices in fixed-effects models when the variability over time is influenced by the baseline covariate profile. Pourahmadi and Daniels (2002) extended the modified Cholesky decomposition for covariance matrices in fixed-effects models to conditionally linear mixed models. Pan and MacKenzie (2007) utilized the modified Cholesky decomposition in linear mixed-effects models.

In clinical trials, unbalanced longitudinal data are common. However, little research has been done for modeling covariance matrices for unbalanced longitudinal data. Pourahmadi's (1999, 2000) modified Cholesky decomposition works well for balanced longitudinal data. However, this approach is not directly applicable when the longitudinal data are unbalanced, because a Cholesky factorization for observed data that is coherent across all subjects usually does not exist.

In this dissertation, we investigate the modified Cholesky decomposition of covariance matrices in fixed-effects models, and mixed-effects models with unbalanced longitudinal data. We address this concern by treating the problem as a missing data problem, employing a generalized EM algorithm to compute the maximum likelihood (ML) estimates. First, a literature review is presented in Chapter II. Chapter III develops a method for modeling covariance matrices in fixed-effects models with unbalanced longitudinal data. Chapter IV provides a method for modeling covariance matrices in mixed-effects models with unbalanced longitudinal data. Chapter V provides a summary and discussion of future research.

CHAPTER II

A REVIEW OF MODELING COVARIANCE MATRICES WITH MODIFIED CHOLESKY DECOMPOSITION

2.1 Modified Cholesky Decomposition

Modified Cholesky decomposition (Pourahmadi, 1999) is used to transform the constrained covariance matrix to unconstrained new parameters. Then we can use generalized regression to model the covariance matrix, as is done for the mean vector.

Newton (1988, P.359) showed a key result that a symmetric matrix Σ is positive definite if and only if there exists a unique lower triangular matrix T with 1's as its diagonal elements and a unique diagonal matrix D with positive diagonal elements such that

$$T\Sigma T' = D.$$

The below-diagonal entries of T are negatives of the successive regression coefficients, and the diagonal entries of D are the prediction error variances when measurements are regressed on their predecessors. Let $y = (y_1, \ldots, y_m)'$ be a time-ordered random vector with mean zero and covariance matrix Σ. For $1 \leq t \leq m$, let $\hat{y}_t$ stand for the linear least-squares predictor of y_t based on its predecessors $y_{t-1}, \ldots, y_1$, and let $\epsilon_t = y_t - \hat{y}_t$ be its prediction error with variance $\sigma_t^2 = \text{var}(\epsilon_t)$. Thus, for $t = 1$, $\hat{y}_1 = E(y_1) = 0$, and, for $1 < t \leq m$, there are unique scalars ϕ_{tj} such that

$$y_t = \sum_{j=1}^{t-1} \phi_{tj} y_j + \epsilon_t.$$

The prediction errors

$$\epsilon_t = y_t - \sum_{j=1}^{t-1} \phi_{tj} y_j \qquad (t = 1, \ldots, m)$$

are uncorrelated. Let $\epsilon = (\epsilon_1, \ldots, \epsilon_m)'$ be the vector of successive prediction errors. Writing in matrix form we have

$$\epsilon = TY,$$

where T is a unit lower triangular matrix with $-\phi_{t,j}$ in the (t, j)-th position for $2 \leq t \leq m$ and $j = 1, 2, \ldots, t-1$. It follows that

$$\text{cov}(\epsilon) = T\text{cov}(Y)T' = T\Sigma T' = D,$$

so that the matrix T diagonalizes the covariance matrix Σ.

Modified Cholesky decomposition transforms the constrained parameters of Σ to unconstrained and interpretable parameters $\phi_{t,j}$, $\log \sigma_t^2$, for $1 \leq t \leq m$ and $1 \leq j \leq t-1$. We call the new parameters $\phi_{t,j}$ and $\log \sigma_t^2$ as the generalized autoregressive parameters and the log-innovation variances of Σ or Y. Since $\phi_{t,j}$ and $\log \sigma_t^2$ are unconstrained, we can model them with covariates.

2.2 Fixed-Effects Models: Balanced Longitudinal Data

Pourahmadi (1999, 2000) proposed the modified Cholesky decomposition of covariance matrices in fixed-effects models for longitudinal data.

Let y_i be a $m \times 1$ vector containing the responses for subject i, where $i = 1, \ldots, n$. The y_i are assumed to follow the model

$$y_i = X_i \beta + e_i,$$

where X_i is a $m \times p$ known design matrix, β is a $p \times 1$ vector of unknown regression parameters, and the e_i are independently distributed as $N(0, \Sigma)$. Further assume that

e_i is a $m \times 1$ vector that corresponds to the same set of m observation times $t_1, \ldots, t_m$, for all i. Let $T\Sigma T' = D$ be the modified Cholesky decomposition of Σ, where T is a unique lower triangular matrix with 1's as main diagonal entries and D is a unique diagonal matrix with positive diagonal entries. The parameters ϕ_{tj} and $\log \sigma_t^2$ are unconstrained and are modeled as $\phi_{tj} = z_{tj}'\gamma$ and $\log(\sigma_t^2) = u_t'\lambda$, $1 \leq j \leq t-1$, $1 \leq t \leq m$, where z_{tj} and u_t are covariates for covariance matrices and $\gamma \in \Sigma^q$ and $\lambda \in \Sigma^r$ are corresponding regression parameters of interest. The quadratic form Q in the likelihood function can be written as

$$\begin{aligned} Q &= \sum_{i=1}^{n} (y_i - X_i\beta)'\Sigma^{-1}(y_i - X_i\beta) = \sum_{i=1}^{n} e_i' T' D^{-1} T e_i \\ &= \sum_{i=1}^{n} (e_i - \tilde{e}_i)' D^{-1} (e_i - \tilde{e}_i) = \sum_{i=1}^{n}\sum_{t=1}^{m} \frac{(e_{i,t} - \tilde{e}_{i,t})^2}{\sigma_t^2} = \sum_{t=1}^{m} \frac{RSS_t}{\sigma_t^2}, \end{aligned}$$

where

$$RSS_t = \sum_{i=1}^{n} (e_{it} - \sum_{j=1}^{t-1} \phi_{tj} e_{ij})^2 = \sum_{i=1}^{n} \{e_{it} - (\sum_{j=1}^{t-1} z_{tj}' e_{ij})\gamma\}^2 = \sum_{i=1}^{n} \{e_{it} - z'(i,t)\gamma\}^2.$$

Therefore

$$Q = \sum_{i=1}^{n}\sum_{t=1}^{m} \sigma_t^{-2} \{e_{it} - z'(i,t)\gamma\}^2 = \sum_{i=1}^{n} \{e_i - Z(i)\gamma\}' D^{-1} \{e_i - Z(i)\gamma\},$$

where

$$z(i,t) = \sum_{j=1}^{t-1} e_{ij} z_{tj}, \qquad Z(i) = (z(i,1), z(i,2), \ldots, z(i,m))'$$

are $q \times 1$ and $m \times q$ matrices respectively. Minus twice the log likelihood function, except for a constant, has three representations:

$$\begin{aligned} -2L(\beta, \lambda, \gamma; y) &= n \log |\Sigma| + \sum_{i=1}^{n} (y_i - X_i\beta)'\Sigma^{-1}(y_i - X_i\beta) \\ &= n \sum_{t=1}^{m} \log \sigma_t^2 + \sum_{t=1}^{m} \frac{RSS_t}{\sigma_t^2} \\ &= n \sum_{t=1}^{m} \log \sigma_t^2 + \sum_{i=1}^{n} \{e_i - Z(i)\gamma\}' D^{-1} \{e_i - Z(i)\gamma\}. \end{aligned}$$

The score function and the Fisher information for the parameters are computed as follows:

$$U_1(\beta) = \sum_{i=1}^{n} X_i' \Sigma^{-1} e_i, \quad I_{11} = \sum_{i=1}^{n} X_i' \Sigma^{-1} X_i,$$
$$U_2(\lambda) = \frac{1}{2} U'(D^{-1}R - n1_m), \quad I_{22} = \frac{1}{2} nU'U,$$
$$U_3(\gamma) = \sum_{i=1}^{n} Z'(i) D^{-1} \{e_i - Z(i)\gamma\}, \quad I_{33} = nW,$$

where $R = (Te')^{(2)} 1_m$ with $e = (e_1', \ldots, e_n')'$ and $1_m = (1, \ldots, 1)'$,

$$U = (u_1, \ldots, u_m)',$$

$$W = E\{Z'(i) D^{-1} Z(i)\} = \sum_{t=1}^{m} \sigma_t^{-2} E\{z(i,t) z'(i,t)\} = \sum_{t=1}^{m} \sigma_t^{-2} W_t,$$

and

$$W_t = E\{z(i,t) z'(i,t)\} = \sum_{k=1}^{t-1} \sum_{l=1}^{t-1} E(e_{ik} e_{il}) z_{tk} z_{tl}' = \sum_{k=1}^{t-1} \sum_{l=1}^{t-1} \sigma_{kl} z_{tk} z_{tl}'.$$

Since $E(e_i) = 0$,

$$I_{12} = E(-\frac{\partial U_1}{\partial \lambda}) = 0, \qquad I_{13} = E(-\frac{\partial U_1}{\partial \gamma}) = 0.$$

After some algebra,

$$I_{23} = nU' D^{-1} B,$$

where

$$b_t = \sum_{j=1}^{t-1} a_{jt} z_{tj}, \qquad B = (b_1, \ldots, b_m)'$$

and a_{jt} is the (j,t)-th entry of the matrix $A = \Sigma T'$.

The maximum likelihood estimates of β and γ can be obtained by setting the score function U_1 and U_3 to zero

$$\tilde{\beta} = \tilde{\beta}(\Sigma) = (\sum_{i=1}^{n} X_i' \Sigma^{-1} X_i)^{-1} \sum_{i=1}^{n} X_i' \Sigma^{-1} y_i,$$

$$\tilde{\gamma} = \tilde{\gamma}(\beta, D) = \{\sum_{i=1}^{n} Z'(i)D^{-1}Z(i)\}^{-1} \sum_{i=1}^{n} Z'(i)D^{-1}e_i.$$

An iterative Fisher scoring method for obtaining $\tilde{\beta}, \tilde{\lambda}, \tilde{\gamma}$ (Pourahmadi, 2000) is provided as follows:

1. Initialization: set β as its ordinary least-squares estimate.

2. Update the estimated covariance matrix $S = n^{-1} \sum_{i=1}^{n} (y_i - X_i\beta)(y_i - X_i\beta)'$ to obtain the estimates of T and D.

3. Update λ, γ using the Newton-Raphson iterative method with Fisher scoring. Update T, D and $\Sigma^{-1} = T'D^{-1}T$ at convergence.

4. Update β using

$$\tilde{\beta} = (\sum_{i=1}^{n} X_i'\tilde{\Sigma}^{-1}X_i)^{-1} \sum_{i=1}^{n} X_i'\tilde{\Sigma}^{-1}y_i.$$

5. Iterate 2-4 until convergence.

2.3 Fixed-Effects Models: Unbalanced Longitudinal Data

Pan and MacKenzie (2003) suggested a more general algorithm which takes unbalanced data into account.

Let y_{it} be the t-th of m_i measurements on the i-th of n subjects and let t_{it} be the time at which the measurement y_{it} is made. Denote by $y_i = (y_{i1}, y_{i2}, \ldots, y_{im_i})'$ and $t_i = (t_{i1}, t_{i2}, \ldots, t_{im_i})'$ the $m_i \times 1$ vectors of responses and times of the ith subject. It is assumed that y_i are independently distributed as $N_{m_i}(\mu_i, \Sigma_i)$, where $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{im_i})'$ is an $m_i \times 1$ vector and Σ_i is an $m_i \times m_i$ positive definite matrix. The mean μ_{it} of y_{it} are assumed to follow the model

$$\mu_{it} = x_{it}'\beta,$$

where x_{it} is the baseline covariates associated with the t-th observation of the i-th subject, and β is a $p \times 1$ vector of unknown regression parameters. The subject-specific

covariance matrix, Σ_i may be modeled using modified Cholesky decomposition. Let $T_i\Sigma_i T_i' = D_i$ be the modified Cholesky decomposition of Σ_i, where T_i is a unique lower triangular matrix with 1's as main diagonal entries and D_i is a unique diagonal matrix with positive diagonal entries. The parameters ϕ_{itj} and $\log\sigma_{it}^2$ are unconstrained and are modeled as $\phi_{itj} = z_{itj}'\gamma$ and $\log(\sigma_{it}^2) = u_{it}'\lambda$, $1 \leq j \leq t-1$, $1 \leq t \leq m_i$, where z_{itj} and u_{it} are covariates for covariance matrices and $\gamma \in R^q$ and $\lambda \in R^r$ are corresponding regression parameters of interest. Minus twice the log likelihood function, except for a constant, is given by

$$-2L = \sum_{i=1}^{n} \log|T_i^{-1}D_iT_i'^{-1}| + \sum_{i=1}^{n} r_i'T_i'D_i^{-1}T_ir_i,$$

where $r_{it} = y_{it} - x_{it}'\beta$ is the t-th element of $r_i = y_i - X_i\beta$, the vector of residuals, and the matrix X_i has row vectors $x_{it}'(t = 1, 2, \ldots, m_i)$. The maximum likelihood estimating equations for β, γ, λ become

$$U_1(\beta) = \sum_{i=1}^{n} X_i'\Sigma_i^{-1}(y_i - X_i\beta) = 0,$$
$$U_2(\gamma) = \sum_{i=1}^{n} Z_i^{*'}D_i^{-1}(r_i - Z_i^*\gamma) = 0,$$
$$U_3(\lambda) = \frac{1}{2}\sum_{i=1}^{n} U_i'(D_i^{-1}e_i - 1_{m_i}) = 0,$$

where the matrix Z_i^*, a $m_i \times q$ matrix, has row $z_{it}^{*'} = \sum_{j=1}^{t-1} r_{ij}z_{itj}'$. Also,

$$U_i = (u_{i1}', u_{i2}', \ldots u_{im_i}')', e_i = (e_{i1}', e_{i2}', \ldots, e_{im_i}')',$$

with $e_{it} = (r_{it} - \hat{r}_{it})^2$ and $\hat{r}_{it} = \sum_{j=1}^{t-1}\phi_{itj}r_{ij}$, are the $m_i \times q$ matrix of covariates and the $m_i \times 1$ vector of squared fitted residuals, respectively, and 1_{m_i} is the $m_i \times 1$ vector of $1's$.

An iteratively re-weighted least squares algorithm is used to calculate the maximum likelihood estimates.

1. Initialization: set $\Sigma_i = I_{m_i} (i = 1, 2, \ldots, n)$.

2. Using the current estimates of γ and λ, update β using

$$\tilde{\beta} = (\sum_{i=1}^{n} X_i' \Sigma_i^{-1} X_i)^{-1} \sum_{i=1}^{n} X_i' \Sigma_i^{-1} y_i.$$

3. Using the current estimates of β and λ, update γ by

$$\tilde{\gamma} = \{\sum_{i=1}^{n} Z_i^{*'} D_i^{-1} Z_i^*\}^{-1} \sum_{i=1}^{n} Z_i^{*'} D_i^{-1} r_i,$$

where Z_i^* is a $m_i \times q$ matrix.

4. Using the current estimates of β and γ, update λ using the Newton-Raphson iterative method with Fisher scoring.

5. Iterate 2-4 until convergence.

This is a more general approach than that given by Pourahmadi (1999). Pan and MacKenzie's algorithm can be applied to the dropout longitudinal data. However, this approach is not directly applicable when the longitudinal data have intermittent missing values, because a Cholesky factorization for observed data coherent across all subjects usually does not exist.

2.4 Including Baseline Covariates in Modeling Covariance Matrices

Pan and MacKenzie (2006) considered including baseline covariates in modeling covariance structures under the circumstances that the variability over time is influenced by the baseline covariate profile. Then, the parameters ϕ_{itj} and $\log \sigma_{it}^2$ are modeled as $\phi_{itj} = \tilde{x}_{itj}' \beta_2 + z_{itj}' \gamma$ and $\log(\sigma_{it}^2) = \tilde{x}_{it}' \beta_3 + u_{it}' \lambda$, $1 \leq j \leq t-1$, $1 \leq t \leq m_i$, where z_{itj} and u_{it} are covariates for covariance matrices, $\gamma \in \Sigma^q$ and $\lambda \in \Sigma^r$ are corresponding regression parameters of interest, $\tilde{x}_{itj}$ and $\tilde{x}_{it}$ are baseline covariates for the ith subject and interaction terms with z_{itj} and u_{it} respectively. The intercept terms are usually included in the z_{itj} and u_{it} when polynomial terms are fitted. If two distinct

polynomials with $q-1$ and $d-1$ degrees are used to model the correlation parameters ϕ_{itj} and log-innovation variances $\log \sigma_{it}^2$, the z_{itj} and u_{it} can have the forms

$$z_{itj} = (1, (t_{it} - t_{ij}), (t_{it} - t_{ij})^2, \ldots, (t_{it} - t_{ij})^{q-1})',$$
$$u_{it} = (1, t_{it}, t_{it}^2, \ldots, t_{it}^{d-1})'.$$

Given one baseline covariate a_i which indicates treatment A when $a_i = 0$ and treatment B when $a_i = 1$, the covariates $\tilde{x}_{itj}$ and $\tilde{x}_{it}$ may have the forms

$$\tilde{x}_{itj} = (a_i, a_i \times (t_{it} - t_{ij}), a_i \times (t_{it} - t_{ij})^2, \ldots, a_i \times (t_{it} - t_{ij})^{q-1})',$$
$$\tilde{x}_{it} = (a_i, a_i \times t_{it}, a_i \times t_{it}^2, \ldots, a_i \times t_{it}^{d-1})'.$$

Minus twice the log-likelihood function, except for a constant, has three representations:

$$\begin{aligned} -2L(\beta, \lambda, \gamma; y) &= \sum_{i=1}^{n} \log |\Sigma_i| + \sum_{i=1}^{n} (y_i - X_i\beta)' \Sigma_i^{-1} (y_i - X_i\beta) \\ &= \sum_{i=1}^{n} \sum_{t=1}^{m_i} \log \sigma_{it}^2 + \sum_{i=1}^{n} \sum_{t=1}^{m_i} \frac{(r_{it} - \tilde{r}_{it})^2}{\sigma_{it}^2} \\ &= \sum_{i=1}^{n} \log |D_i| + \sum_{i=1}^{n} (r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma)' D_i^{-1} (r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma), \end{aligned}$$

where $r_{it} = y_{it} - x_{it}'\beta$ is the t-th element in the vector $r_i = y_i - x_i\beta$ of order $m_i \times 1$, $\tilde{r}_{it} = \sum_{j=1}^{t-1} \phi_{itj} r_{ij}$ is the t-th element in the vector $\tilde{r}_i$ of order $m_i \times 1$, the t-th row in $\tilde{X}_i^*$ is $\tilde{x}_{it}^{*\prime}$ with $\tilde{x}_{it}^* = \sum_{j=1}^{t-1} r_{it} \tilde{x}_{itj}$, the t-th row of Z_i^* is $z_{it}^{*\prime}$ with $z_{it}^* = \sum_{j=1}^{t-1} r_{it} z_{itj}$. Note that $\tilde{x}_{i1}^* = 0$, $z_{i1}^* = 0$ and $\tilde{r}_{i1} = 0$.

The score functions for the parameters are:

$$U_1(\beta) = \sum_{i=1}^{n} X_i' \Sigma_i^{-1} (y_i - X_i \beta),$$
$$U_2(\beta_2) = \sum_{i=1}^{n} \tilde{X}_i^{*'} D_i^{-1} (r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma),$$
$$U_3(\gamma) = \sum_{i=1}^{n} Z_i^{*'} D_i^{-1} (r_i - \tilde{X}_i^* \beta_2 - Z_i^* \gamma),$$
$$U_4(\beta_3) = \frac{1}{2} \sum_{i=1}^{n} \tilde{X}_i' (D_i^{-1} e_i - 1_{m_i}),$$
$$U_5(\lambda) = \frac{1}{2} \sum_{i=1}^{n} U_i' (D_i^{-1} e_i - 1_{m_i}),$$

where $e_i = (e_{i1}, e_{i2}, \ldots, e_{im_i})'$ with $e_{it} = (r_{it} - \tilde{r}_{it})^2$, $\tilde{X}_i = (\tilde{X}_{i1}', \tilde{X}_{i2}', \ldots, \tilde{X}_{im_i}')'$ and $U_i = (u_{i1}', u_{i2}', \ldots, u_{im_i}')'$.

The maximum likelihood estimates of β, β_2 and γ can be obtained by setting the score functions U_1, U_2 and U_3 to zero. Then we get

$$\tilde{\beta} = (\sum_{i=1}^{n} X_i' \Sigma_i^{-1} X_i)^{-1} \sum_{i=1}^{n} X_i' \Sigma_i^{-1} y_i,$$

$$\begin{pmatrix} \tilde{\beta}_2 \\ \tilde{\gamma} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \tilde{X}_i^{*'} D_i^{-1} \tilde{X}_i^* & \sum_{i=1}^{n} \tilde{X}_i^{*'} D_i^{-1} Z_i^* \\ \sum_{i=1}^{n} Z_i^{*'} D_i^{-1} \tilde{X}_i^* & \sum_{i=1}^{n} Z_i^{*'} D_i^{-1} Z_i^* \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} \tilde{X}_i^{*'} D_i^{-1} r_i \\ \sum_{i=1}^{n} Z_i^{*'} D_i^{-1} r_i \end{pmatrix}.$$

The Newton-Raphson algorithm is applied to compute the estimates of β_3 and λ. The updated estimates of β_3 and λ are

$$\begin{pmatrix} \tilde{\beta}_3 \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \tilde{X}_i' V_i^{-1} \tilde{X}_i & \sum_{i=1}^{n} \tilde{X}_i' V_i^{-1} U_i \\ \sum_{i=1}^{n} U_i' V_i^{-1} \tilde{X}_i & \sum_{i=1}^{n} U_i' V_i^{-1} U_i \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} \tilde{X}_i' V_i^{-1} \mu_i \\ \sum_{i=1}^{n} U_i' V_i^{-1} \mu_i \end{pmatrix},$$

where $\mu_i = (\log D_i + I_{m_i} - V_i) 1_{m_i}$ and $V_i^{-1} = diag(e_{i1}/\sigma_{i1}^2, \ldots, e_{im_i}/\sigma_{im_i}^2)$.

To calculate the maximum likelihood estimates of the parameters, Pan and MacKenzie (2006) proposed the following algorithm:

1. Initialization: set $\Sigma_i = I_{m_i} (i = 1, 2, \ldots, n)$.

2. Using the current estimates of β_2, β_3, γ and λ, update β using

$$\tilde{\beta} = (\sum_{i=1}^{n} X_i' \tilde{\Sigma}_i^{-1} X_i)^{-1} \sum_{i=1}^{n} X_i' \tilde{\Sigma}_i^{-1} y_i.$$

3. Using the current estimates of β, β_3 and λ, update β_2 and γ.

4. Using the current estimates of β, β_2 and γ, update λ and β_3 using the Newton-Raphson iterative method.

5. Iterate 2-4 until convergence.

2.5 Conditionally Linear Mixed Models

Pourahmadi and Daniels (2002) extended the modified Cholesky decomposition for covariance matrices in fixed-effects models to conditionally linear mixed models for longitudinal data. The so-called dynamic conditionally linear mixed model has the form

$$y_i = f(X_i \beta) + Z_i b_i + \epsilon_i^*,$$

where y_i is an $m_i \times 1$ vector containing the responses for subject i, $i = 1, \ldots, n$, X_i is a $m_i \times p$ design matrix, β is a $p \times 1$ vector of unknown fixed effects parameters, Z_i is an $m_i \times q$ design matrix, b_i is a $q \times 1$ vector of unknow random effects parameters, f is a known nonlinear function of its arguments (Vonesh and Carter, 1992), $\epsilon_i^* \sim N(0, \Sigma_i)$ and $\Sigma_i^{-1} = T' D_i^{-1} T$.

The linear parameters in the model may be random, but nonlinear parameters are fixed. This setup allows for flexible models and includes many covariance structures such as all standard linear mixed models, ante-dependence models and Vonesh-Carter models. It guarantees the fitted covariance matrix is positive definite.

2.6 Linear Mixed-Effects Models: Balanced Longitudinal Data

Pan and MacKenzie (2007) suggested modeling the within subject covariance matrices in linear mixed models with modified Cholesky decomposition (Pourahmadi, 1999, 2000).

The linear mixed model has the form

$$y_i = X_i\beta + Z_i\alpha_i + e_i,$$

where y_i is an $m_i \times 1$ vector containing the responses for subject i, $i = 1, \ldots, n$, X_i is a $m_i \times p$ known matrix, β is a $p \times 1$ vector of unknown regression parameters, Z_i is an $m_i \times b$ design matrix, the $b \times 1$ vector of between-subject random effects α_i are distributed as $N(0, B)$, and the $m_i \times 1$ vector of within-subject random effects e_i are independently distributed as $N(0, \Sigma_i)$. The subject specific covariance matrix Σ_i can be modeled using covariates following the approach of Pourahmadi (1999, 2000). Let $T_i\Sigma_i T_i' = D_i$ be the modified Cholesky decomposition of Σ_i, where T_i is a unique lower triangular matrix with 1's as main diagonal entries and D_i is a unique diagonal matrix with positive diagonal entries. The parameters ϕ_{itj} and $\log \sigma_{it}^2$ are unconstrained and are modeled as $\phi_{itj} = z_{itj}'\gamma$ and $\log(\sigma_{it}^2) = u_{it}'\lambda$, $1 \leq j \leq t-1$, $1 \leq t \leq m$, where z_{itj} and u_{it} are covariates for covariance matrices and $\gamma \in R^q$ and $\lambda \in R^r$ are corresponding regression parameters of interest. The between-subject covariance matrix is denoted by the general parameter vector δ, i.e., $B = B(\delta)$.

The EM-algorithm is adopted to calculate the maximum likelihood estimates of the parameters $\theta = (\beta', \alpha', \gamma', \lambda')$. The logarithm of the joint density function of the

responses and random effects is given by

$$\log f(y,\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\log|\Sigma_i| - \frac{1}{2}\sum_{i=1}^{n}(y_i - X_i\beta - Z_i\alpha_i)'\Sigma_i^{-1}(y_i - X_i\beta - Z_i\alpha_i) - \frac{n}{2}\log|B| - \frac{1}{2}\sum_{i=1}^{n}\alpha_i' B^{-1}\alpha_i.$$

If the random effects are treated as missing data, let Q be the expected likelihood given the observed data y and the current parameter values $\theta = (\beta^{(s)'}, \alpha^{(s)'}, \gamma^{(s)'}, \lambda^{(s)'})$. Then

$$\begin{aligned} Q &= E\{\log f(y,\alpha)|y,\theta^{(s)}\} \\ &= -\frac{1}{2}\sum_{i=1}^{n}\log|\Sigma_i| - \frac{1}{2}\sum_{i=1}^{n}\widehat{e}_i^{(s)'}\Sigma_i^{-1}\widehat{e}_i^{(s)} - \frac{1}{2}\sum_{i=1}^{n}tr\{\Sigma_i^{-1}Z_i(\Phi_i^{(s)})^{-1}Z_i'\} \\ &\quad - \frac{n}{2}\log|B| - \frac{1}{2}\sum_{i=1}^{n}tr[B^{-1}\{(\Phi_i^{(s)})^{-1} + \widehat{\alpha}_i^{(s)}\widehat{\alpha}_i^{(s)'}\}], \end{aligned}$$

where

$$\begin{aligned} \widehat{e}_i^{(s)} &= y_i - X_i\beta - Z_i\widehat{\alpha}_i^{(s)}, \\ \widehat{\alpha}_i^{(s)} &= B^{(s)}Z_i'(Z_iB^{(s)}Z_i' + \Sigma_i^{(s)})^{-1}(y_i - X_i\beta^{(s)}), \\ \Phi_i^{(s)} &= (B^{(s)})^{-1} + Z_i'(\Sigma_i^{(s)})^{-1}Z_i. \end{aligned}$$

For the fixed effects, the first-derivative of Q with respect to β betted to zero leads to the updated β

$$\beta^{(s+1)} = \{\sum_{i=1}^{n}X_i'(\Sigma_i^{(s)})^{-1}X_i\}^{-1}\sum_{i=1}^{n}X_i'(\Sigma_i^{(s)})^{-1}(y_i - Z_i\widehat{\alpha}_i^{(s)}).$$

The estimation for the between-subject covariance matrix B is obtained by taking first-derivative of Q with respect to δ_k, the k-th component of the vector δ, and setting to zero. We have

$$\frac{\partial Q}{\partial\delta_k} = -\frac{1}{2}tr\Big([nB^{-1} - B^{-1}\sum_{i=1}^{n}\{\widehat{\alpha}_i^{(s)}\widehat{\alpha}_i^{(s)'} + (\Phi_i^{(s)})^{-1}\}B^{-1}]\frac{\partial B}{\partial\delta_k}\Big) = 0.$$

Then the updated between-subject covariance matrix B is

$$B^{(s+1)} = \frac{1}{n}\sum_{i=1}^{n}\{\alpha_i^{(s)}\alpha_i^{(s)'} + (\Phi_i^{(s)})^{-1}\}.$$

Newton-Raphson algorithm is used to update γ and λ. The first-derivatives of Q with respect to γ and λ are given by

$$\frac{\partial Q}{\partial \gamma} = \sum_{i=1}^{n}\sum_{t=1}^{m_i}\sigma_{it}^{-2}[A_{it}^{\hat{e}^{(s)}}\{\widehat{e}_{it}^{(s)} - (A_{it}^{\hat{e}^{(s)}})'\gamma\} + A_{it}^{\tilde{z}^{(s)}}\{\tilde{z}_{it}^{(s)} - (A_{it}^{\tilde{z}^{(s)}})'\gamma\}]$$

and

$$\begin{aligned}\frac{\partial Q}{\partial \lambda} = &-\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m_i}h_{it} + \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m_i}\sigma_{it}^{-2}\{\widehat{e}_{it}^{(s)} - (A_{it}^{\hat{e}^{(s)}})'\gamma\}^2 h_{it} \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m_i}\sigma_{it}^{-2}\{\tilde{z}_{it}^{(s)} - (A_{it}^{\tilde{z}^{(s)}})'\gamma\}'\{\tilde{z}_{it}^{(s)} - (A_{it}^{\tilde{z}^{(s)}})'\gamma\}h_{it},\end{aligned}$$

where $A_{it}^{\hat{e}^{(s)}} = \sum_{k=1}^{t-1} z_{itk}\widehat{e}_{ik}^{(s)}$ and $\widehat{e}_{it}^{(s)}$ are the transposes of t-th rows of $\widehat{e}_i$, $A_{it}^{\tilde{z}^{(s)}} = \sum_{k=1}^{t-1} z_{itk}(\tilde{z}_{ik}^{(s)})'$ and $\tilde{z}_{it}^{(s)}$ are the transposes of the t-th rows of $\tilde{Z}_i^{(s)} = Z_i(\Phi_i^{(s)})^{-1/2}$. When $t = 1$, we have $A_{i1}^{\widehat{e}^{(s)}} = 0$ and $A_{i1}^{\tilde{z}^{(s)}} = 0$. The second-derivatives of Q with respect to γ and λ are given by

$$\begin{aligned}\frac{\partial^2 Q}{\partial\gamma\partial\gamma'} &= -\sum_{i=1}^{n}\sum_{t=1}^{m_i}\sigma_{it}^{-2}\{A_{it}^{\hat{e}^{(s)}}(A_{it}^{\hat{e}^{(s)}})' + A_{it}^{\tilde{z}^{(s)}}(A_{it}^{\tilde{z}^{(s)}})'\}, \\ \frac{\partial^2 Q}{\partial\lambda\partial\lambda'} &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m_i}\sigma_{it}^{-2}[\{\widehat{e}_{it}^{(s)} - (A_{it}^{\hat{e}^{(s)}})'\gamma\}^2 + \{\tilde{z}_{it}^{(s)} - (A_{it}^{\tilde{z}^{(s)}})'\gamma\}'\{\tilde{z}_{it}^{(s)} - (A_{it}^{\tilde{z}^{(s)}})'\gamma\}](h_{it}h_{it}'), \\ \frac{\partial^2 Q}{\partial\gamma\partial\lambda'} &= -\sum_{i=1}^{n}\sum_{t=1}^{m_i}\sigma_{it}^{-2}[A_{it}^{\hat{e}^{(s)}}\{\widehat{e}_{it}^{(s)} - (A_{it}^{\hat{e}^{(s)}})'\gamma\} + A_{it}^{\tilde{z}^{(s)}}\{\tilde{z}_{it}^{(s)} - (A_{it}^{\tilde{z}^{(s)}})'\gamma\}]h_{it}'.\end{aligned}$$

Hence the updated γ and λ using the Newton-Raphson algorithm are

$$\begin{pmatrix}\gamma^{(s+1)} \\ \lambda^{(s+1)}\end{pmatrix} = \begin{pmatrix}\gamma^{(s)} \\ \lambda^{(s)}\end{pmatrix} + \begin{pmatrix}-\frac{\partial^2 Q}{\partial\gamma\partial\gamma'} & -\frac{\partial^2 Q}{\partial\gamma\partial\lambda'} \\ -\frac{\partial^2 Q}{\partial\lambda\partial\gamma'} & -\frac{\partial^2 Q}{\partial\lambda\partial\lambda'}\end{pmatrix}^{-1}_{\theta=\theta^{(s)}} \begin{pmatrix}\frac{\partial Q}{\partial\gamma} \\ \frac{\partial Q}{\partial\lambda}\end{pmatrix}_{\theta=\theta^{(s)}}.$$

The steps of the EM-algorithm for the estimation of the parameters are

1. E-step: update the random effects $\alpha_i^{(s)}$ given current values of β, δ, γ and λ. Then update the between-subject covariance matrix B.

2. M-step: update the fixed effects β. Then update the γ and λ via the Newton-Raphson algorithm.

3. Iterate the E-step and M-step until convergence.

CHAPTER III

FIXED-EFFECTS MODELS WITH UNBALANCED LONGITUDINAL DATA

3.1 Introduction

Pourahmadi's (1999, 2000) modified Cholesky decomposition works well for balanced longitudinal data. However, this approach is not directly applicable when the longitudinal data are unbalanced, because a Cholesky factorization for observed data coherent across all subjects usually does not exist. In this chapter, we address this concern by treating the problem as a missing data problem and employing a generalized EM algorithm to compute the ML estimators for the fixed-effects models. Section 3.2 presents the incomplete data fixed-effects model and details for the modified Cholesky decomposition of the covariance matrix. The algorithm to calculate the ML estimators is given in Section 3.3. Section 3.4 provides the asymptotic distribution of the ML estimators. Section 3.5 and 3.6 describe an application to Kenward's (1987) cattle data and simulation studies.

3.2 The Incomplete Data Model

Let y_i be a $m_i \times 1$ vector containing the responses for subject i, where $i = 1, \ldots, n$. The y_i are assumed to follow the model

$$y_i = X_i\beta + e_i,$$

where X_i is a $m_i \times p$ known matrix, β is a $p \times 1$ vector of unknown regression parameters, and the e_i are independently distributed as $N(0, \Sigma_i)$. We further assume that e_i is a sub-vector of a $m \times 1$ vector e_i^* that corresponds to the same set of m observation times $t_1, \ldots, t_m$, for all i. It follows that Σ_i is a sub-matrix of $\Sigma_i^* = \text{var}(e_i^*)$.

This model, called the incomplete data model, arises in situations where a fixed number m of measurements, corresponding to different times, are to be collected on each of n subjects, but not all of the subjects' responses are observed (Jennrich and Schluchter, 1986).

The subject specific covariance matrix Σ_i^* can be modeled using covariates following the approach of Pourahmadi (1999, 2000). There exists a unique lower triangular matrix T_i with 1's as main diagonal entries and a unique diagonal matrix D_i with positive diagonal entries such that $T_i\Sigma_i^*T_i' = D_i$. The below-diagonal entries of T_i are the negatives of the autoregressive coefficients, ϕ_{itj}, in $\tilde{e}_{it}^* = \sum_{j=1}^{t-1}\phi_{itj}e_{ij}^*$, the linear least squares predictor of e_{ij}^* based on its predecessors $e_{i(t-1)}^*, \ldots, e_{i1}^*$. The diagonal entries of D_i are the innovation variances $\sigma_{it}^2 = \text{var}(e_{it}^* - \tilde{e}_{it}^*)$, where $1 \leq t \leq m$ and $1 \leq i \leq n$. The parameters ϕ_{itj} and $\log\sigma_{it}^2$ are unconstrained and are modeled as $\phi_{itj} = z_{itj}'\gamma$ and $\log(\sigma_{it}^2) = u_{it}'\lambda$, $1 \leq j \leq t-1$, $1 \leq t \leq m$, where z_{itj} and u_{it} are covariates for covariance matrices and $\gamma \in R^q$ and $\lambda \in R^r$ are corresponding regression parameters of interest. We assume that there is no missing value in these covariates, which is the case if they only depend on baseline covariates and scheduled observation times.

The model is flexible. The $\log(\sigma_{it}^2)$ and ϕ_{itj} can have nonlinear functions of their parameters γ, λ respectively. For example, the matrices T with entries $\phi_{i,j}$ given respectively by

$$\gamma^{i-j}, \gamma^{(t_i - t_{i-j})^\theta}, \gamma_j^{f(t_i,\lambda_j)-f(t_{i-j},\lambda_j)}, \gamma_{i-j}$$

are correspond to $AR(1)$, damped exponential (Diggle, 1988; Munoz, Carey, Schouten, Segal, and Rosner, 1992), structured ante-dependence (Zimmerman and Nunez-Anton, 1997) and banded (Pourahmadi, 1999).

3.3 Algorithm for Maximum Likelihood

This section derives a generalized EM algorithm (Dempster, Laird, and Rubin, 1977; Jennrich and Schluchter, 1986; Laird, Lange, and Stram, 1987) to compute the ML estimators. The iterative algorithm consists of two parts. The first part uses generalized least squares to update β:

$$\tilde{\beta} = \Big(\sum_{i=1}^{n} X_i'\Sigma_i^{-1}X_i\Big)^{-1}\Big(\sum_{i=1}^{n} X_i'\Sigma_i^{-1}y_i\Big), \tag{3.1}$$

which is equivalent to maximizing the likelihood with respect to β while holding γ and λ fixed. The second part comprises one iteration of a generalized EM algorithm to update λ and γ, using e_i^* as complete data and sub-vectors $e_i = y_i - X_i\beta$ as observed data, and assuming β is equal to its current value. The updates of γ and λ are obtained separately using weighted least squares and one step of Newton-Raphson.

Now we give details of the second part of the algorithm. Minus twice the log likelihood function for complete data, except for a constant, is given by

$$-2l = \sum_{i=1}^{n}(\log|\Sigma_i^*| + e_i^{*\prime}\Sigma_i^{*-1}e_i^*) = \sum_{i=1}^{n}\{\log|\Sigma_i^*| + \mathrm{tr}(\Sigma_i^{*-1}V_i)\}.$$

where $V_i = e_i^*e_i^{*\prime}$. Let Q be the expected log likelihood given the observed data and the current parameter values. Then

$$-2Q = \sum_{i=1}^{n}\{\log|\Sigma_i^*| + \mathrm{tr}(\Sigma_i^{*-1}\widehat{V}_i)\},$$

where $\widehat{V}_i = E(e_i^*e_i^{*\prime}|e_i) = \widehat{e_i^*}\widehat{e_i^*}' + \mathrm{var}(e_i^*|e_i)$ with $\widehat{e_i^*} = E(e_i^*|e_i)$. The calculation of $\widehat{e_i^*}$ and $\widehat{V}_i$ follows standard results for multivariate normal distributions. For example, if

$$e_i^* = \begin{pmatrix} e_i \\ e_i^+ \end{pmatrix} \sim N(0, \Sigma_i^*), \qquad \Sigma_i^* = \begin{pmatrix} \Sigma_{i11}^* & \Sigma_{i12}^* \\ \Sigma_{i21}^* & \Sigma_{i22}^* \end{pmatrix},$$

then

$$E(e_i^*|e_i) = \begin{pmatrix} I \\ \Sigma_{i21}^* \Sigma_{i11}^{*-1} \end{pmatrix} e_i \tag{3.2}$$

and

$$\mathrm{var}(e_i^*|e_i) = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{i22}^* - \Sigma_{i21}^* \Sigma_{i11}^{*-1} \Sigma_{i12}^* \end{pmatrix}. \tag{3.3}$$

In the EM algorithm we need to minimize $-2Q$ with respect to γ and λ. We now derive two expressions of $-2Q$ that are useful in solving the optimization problem. First note that $|\Sigma_i^*| = |D_i| = \prod_{t=1}^m \sigma_{it}^2$. Since $\Sigma_i^{*-1} = T_i' D_i^{-1} T_i$ and the t-th row of $T_i e_i^*$ is

$$e_{it}^* - \sum_{j=1}^{t-1} \phi_{itj} e_{ij}^* = e_{it}^* - \sum_{j=1}^{t-1} e_{ij}^* z_{itj}' \gamma,$$

we have that

$$e_i^{*'} \Sigma_i^{*-1} e_i^* = e_i^{*'} T_i' D_i^{-1} T_i e_i^* = \sum_{t=1}^m \sigma_{it}^{-2} \mathrm{RS}_{it},$$

where the residual squared $\mathrm{RS}_{it} = (e_{it}^* - \sum_{j=1}^{t-1} e_{ij}^* z_{itj}' \gamma)^2$. Denote $Z_{it}' = (z_{it1}, \ldots, z_{it(t-1)})$ and recall that $V_i = (V_{ijj'}) = e_i^* e_i^{*'}$. Let $V_{it} = V_i[\cdot, t]$, $V_{it}^{(t-1)} = V_i[1{:}(t-1), t]$ and $V_i^{(t-1)} = V_i[1{:}(t-1), 1{:}(t-1)]$ be appropriate sub-matrices of V_i. Then

$$\begin{aligned} \mathrm{RS}_{it} &= e_{it}^{*2} - 2 \sum_{j=1}^{t-1} \gamma' z_{itj} e_{ij}^* e_{it}^* + \gamma' \Big(\sum_{j,j'=1}^{t-1} z_{itj} e_{ij}^* e_{ij'}^* z_{itj'}' \Big) \gamma \\ &= V_{itt} - 2\gamma' Z_{it}' V_{it}^{(t-1)} + \gamma' Z_{it}' V_i^{(t-1)} Z_{it} \gamma. \end{aligned}$$

Note that if $t = 1$ we have $V_{it}^{(t-1)} = 0$ and $V_i^{(t-1)} = 0$. The above discussion yields two expressions of the log likelihood of complete data

$$\begin{aligned} -2l &= \sum_{i=1}^n \sum_{t=1}^m \Big(\log \sigma_{it}^2 + \frac{\mathrm{RS}_{it}}{\sigma_{it}^2} \Big) \\ &= \sum_{i=1}^n \sum_{t=1}^m \Big(\log \sigma_{it}^2 + \frac{V_{itt}}{\sigma_{it}^2} \Big) \\ &\quad + \sum_{i=1}^n \sum_{t=1}^m \sigma_{it}^{-2} (-2\gamma' Z_{it}' V_{it}^{(t-1)} + \gamma' Z_{it}' V_i^{(t-1)} Z_{it} \gamma). \end{aligned}$$

It follows that the expected log likelihood given observed data has two expressions

$$\begin{aligned}-2Q &= \sum_{i=1}^{n}\sum_{t=1}^{m}\Big(\log\sigma_{it}^2 + \frac{\widehat{\mathrm{RS}}_{it}}{\sigma_{it}^2}\Big)\\ &= \sum_{i=1}^{n}\sum_{t=1}^{m}\Big(\log\sigma_{it}^2 + \frac{\widehat{V}_{itt}}{\sigma_{it}^2}\Big)\\ &\quad + \sum_{i=1}^{n}\sum_{t=1}^{m}\sigma_{it}^{-2}(-2\gamma' Z_{it}'\widehat{V}_{it}^{(t-1)} + \gamma' Z_{it}'\widehat{V}_{i}^{(t-1)}Z_{it}\gamma),\end{aligned}$$

where $\widehat{\mathrm{RS}}_{it} = E(\mathrm{RS}_{it}|e_i)$ and $\widehat{V}_i$ with sub- and super-scripts are appropriate sub-matrices of $\widehat{V}_i$. We can also show that $\widehat{\mathrm{RS}}_{it}$ is the (t,t)-th element of the matrix $T_i\widehat{V}_iT_i'$.

The update of γ and λ proceeds as follows. For fixed λ, $-2Q$ is a quadratic form in γ and is minimized by

$$\tilde{\gamma} = \Big(\sum_{i=1}^{n}\sum_{t=1}^{m}\sigma_{it}^{-2}Z_{it}'\widehat{V}_i^{(t-1)}Z_{it}\Big)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\sigma_{it}^{-2}Z_{it}'\widehat{V}_{it}^{(t-1)}. \tag{3.4}$$

For fixed γ, optimization of $-2Q$ over λ does not have a closed-form expression and we rely on the Newton-Raphson algorithm. Since $\log\sigma_{it}^2 = u_{it}'\lambda$, simple calculation yields

$$\frac{\partial Q}{\partial\lambda} = -\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m}\Big(1 - \frac{\widehat{\mathrm{RS}}_{it}}{\sigma_{it}^2}\Big)u_{it}$$

and

$$\frac{\partial^2 Q}{\partial\lambda\partial\lambda'} = -\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{\widehat{\mathrm{RS}}_{it}}{\sigma_{it}^2}u_{it}u_{it}'.$$

The Newton-Raphson algorithm updates the current values $\lambda^{(0)}$ to $\lambda^{(1)}$ using

$$\lambda^{(1)} = \lambda^{(0)} + \Delta\lambda, \qquad \Delta\lambda = -\Big(\frac{\partial^2 Q}{\partial\lambda\partial\lambda'}\Big)^{-1}\frac{\partial Q}{\partial\lambda}. \tag{3.5}$$

For the generalized EM algorithm, we only do one step instead of full iteration of the Newton-Raphson. We need make sure that the log likelihood increases at each step of the generalized EM algorithm, using partial stepping such as step-halving if

necessary. Step-halving works as follows. If $Q(\lambda^{(1)}) \leq Q(\lambda^{(0)})$, we replace $\Delta\lambda$ by its half in the update $\lambda^{(1)} = \lambda^{(0)} + \Delta\lambda$, and continue doing so until $Q(\lambda^{(1)}) > Q(\lambda^{(0)})$.

The steps of the algorithm are summarized as follows:

(i) Initialization: set $\Sigma_i^* = I$, $i = 1, \ldots, n$.

(ii) Using the current estimates of γ and λ (or Σ_i^* in the first iteration), compute the updated estimate $\tilde{\beta}$ of β using equation (3.1).

(iii) Compute $\widehat{V}_i$, $i = 1, \ldots, n$, where the relevant conditional expectations are calculated using (3.2) and (3.3).

(iv) Using the current estimates of β and λ, update γ using (3.4).

(v) Using the current estimates of β and γ, update the current estimate $\lambda^{(0)}$ to $\lambda^{(1)}$ using one step of Newton-Raphson as (3.5). Use step-halving to guarantee that the likelihood is increased.

(vi) Iterate (ii)–(v) until convergence.

3.4 Observed Information Matrix

Since we are in a parametric framework, the standard likelihood theory applies. The ML estimators of β, γ, λ are consistent and normally distributed.

One drawback of the EM algorithm is that it does not automatically provide standard errors for the estimates. If the observed data likelihood $L(\theta; y)$ is available, $\hat{\theta}$ can be obtained, then we can find the standard errors from the observed Fisher information $I(\hat{\theta}; y)$. The asymptotic variance-covariance matrix of the maximum likelihood estimators $\hat{\theta}$ can be computed using the inverse of the observed information matrix $I(\hat{\theta}; y)$, evaluated at the $\hat{\theta}$.

We will follow Oakes (1999) and Pawitan (2001). In general, the conditional

density of the complete data y^* given the observed data y is

$$p_\theta(y^*|y) = \frac{p_\theta(y^*)}{p_\theta(y)}.$$

Therefore,

$$\log L(\theta; y^*|y) = \log L(\theta; y^*) - \log L(\theta; y).$$

Taking the conditional expectation given y gives

$$E\{\log L(\theta; y^*|y)|y, \theta^0\} = Q(\theta|\theta^0) - \log L(\theta; y).$$

Let $h(\theta|\theta^0) = E\{\log L(\theta; y^*|y)|y, \theta^0\}$. So

$$\log L(\theta; y) = Q(\theta|\theta^0) - h(\theta|\theta^0).$$

Taking the derivatives with respect to θ we have

$$S(\theta; y) = \frac{\partial Q(\theta|\theta^0)}{\partial \theta} - \frac{\partial h(\theta|\theta^0)}{\partial \theta}.$$

Now $\partial h(\theta|\theta^0)/\partial\theta = 0$ at $\theta = \theta^0$, since $h(\theta|\theta^0)$ is maximized at $\theta = \theta^0$. So we can express the observed score in terms of the first derivative of Q as

$$S(\theta^0; y) = \frac{\partial Q(\theta|\theta^0)}{\partial \theta}|_{\theta=\theta^0}.$$

This relationship holds at each y_i for independent data $y_1, \ldots, y_n$; therefore

$$S(\theta^0; y_i) = \frac{\partial Q_i(\theta|\theta^0)}{\partial \theta}|_{\theta=\theta^0},$$

where $Q_i = E\{\log L(\theta; y_i^*)|y_i, \theta^0\}$. The estimate of the observed Fisher information is

$$I(\tilde{\theta}; y) \approx \sum_i S(\tilde{\theta}; y_i) S^t(\tilde{\theta}; y_i).$$

We can estimate the variance-covariance matrix of $(\tilde{\gamma}, \tilde{\lambda})$, since the observed Fisher information

$$I(\tilde{\theta}; y) \approx \sum_i S(\tilde{\theta}; y_i) S^t(\tilde{\theta}; y_i) \approx \begin{pmatrix} \sum_i S(\tilde{\gamma}; y_i) S^t(\tilde{\gamma}; y_i) & \sum_i S(\tilde{\gamma}; y_i) S^t(\tilde{\lambda}; y_i) \\ \sum_i S(\tilde{\lambda}; y_i) S^t(\tilde{\gamma}; y_i) & \sum_i S(\tilde{\lambda}; y_i) S^t(\tilde{\lambda}; y_i) \end{pmatrix}, \quad (3.6)$$

where

$$S(\tilde{\theta}; y_i) = [\frac{\partial Q_i(\theta|\theta^{(0)})}{\partial \theta}]_{\theta=\tilde{\theta}.}$$

We have that

$$S(\tilde{\gamma}; y_i) = [\frac{\partial Q_i}{\partial \gamma}]_{\gamma=\tilde{\gamma}} = [\sum_{t=1}^{m} \sigma_{it}^{-2}(Z_{it}'\widehat{V}_{it}^{(t-1)} - Z_{it}'\widehat{V}_i^{(t-1)}Z_{it}\gamma)]|_{\gamma=\tilde{\gamma}}$$

and

$$S(\tilde{\lambda}; y_i) = [\frac{\partial Q_i}{\partial \lambda}]_{\lambda=\tilde{\lambda}} = [-\frac{1}{2}\sum_{t=1}^{m}\Big(1 - \frac{\widehat{\mathrm{RS}}_{it}}{\sigma_{it}^2}\Big)u_{it}]|_{\lambda=\tilde{\lambda}}.$$

Then the asymptotic variances of the ML estimator $\tilde{\theta}$ can be obtained by the inverse of the observed Fisher information (3.6), evaluated at the estimator $\tilde{\theta}$.

Since $\tilde{\beta}$ and $(\tilde{\gamma}, \tilde{\lambda})$ are asymptotically independent and consistent estimators for $(\beta', \gamma', \lambda')'$, the asymptotic covariance matrix of $\tilde{\beta}$ can be estimated by $(\sum_i X_i'\tilde{\Sigma}^{-1}X_i)^{-1}$.

3.5 An Application to Kenward's Cattle Data

Kenward (1987) showed an experiment to study the effect of treatments on intestinal parasites of cattle. Thirty cattle were randomly assigned to each of the treatment groups A and B. The weights of the cattle were measured 11 times over a 133-day period. The intervals between consecutive measures were two weeks, except the last one was one week. The measures were taken at the same set of times and there were no missing observations. The times were rescaled as $t = 1, 2, \ldots, 11$. This is a balanced longitudinal data. The equality of the two within treatment-group covariance matrices was rejected using the classical likelihood ratio test (Zimmerman and Nunez-Anton, 1997). Pourahmadi (1999) used a saturated mean model, that is $\mu = (\mu_1, \ldots, \mu_{11})'$, for the mean response profile and reparameterized the covariance matrix with the modified Cholesky decomposition for the treatment group A. By the regressograms, he suggested that the generalized autoregressive parameters were a

cubic function of the lag and the logarithms of the innovation variances were a cubic function of the time. That is,

$$\tilde{\phi}_{tj} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \epsilon_{tj},$$

and

$$\log \tilde{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \epsilon_t,$$

where $t = 1, 2, \ldots, 11$ and $j = 1, 2, \ldots, t-1$.

Let Poly(r, q) represent polynomial models in t and $t-j$ of degree r for $\log \sigma_t^2$ and degree q for ϕ_{tj}, respectively. Then Poly(r, q) model for Σ has $r + q + 2$ parameters. Pourahmadi (2000) fitted Poly$(3, q)$ models for $q = 0, 1, 2, 3$ to Σ and computed the ML estimators of their parameters, the maximized log-likelihood functions and the corresponding BIC values. Note that BIC is defined as

$$BIC = -\frac{2}{n} L_{max} + p\frac{logn}{n},$$

where n is the sample size, L_{max} is the maximized log-likelihood for the model and p is the number of parameters. The better-fitting models have smaller values of BIC.

Based on the L_{max} and BIC columns in Table 1 (Pourahmadi, 2000), the Poly(3,3) model is preferred.

The objective of our analysis is to apply the modified Cholesky decomposition to unbalanced longitudinal data. We had 10%, 20%, 30%, 40%, or 50% of the cattle data missing completely at random. Then we adopted Pourahmadi's scheme using the Poly(3,3) model and estimated the ML estimators with the generalized EM. We repeated this procedure ten times. The results are shown in Figure 1, Figure 2, Figure 3, Figure 4, and Figure 5. The fitted log-innovation variances and generalized autoregressive parameters from unbalanced data using our generalized EM algorithm are close to those from the complete cattle data using Pourahmadi (1999)'s algorithm.

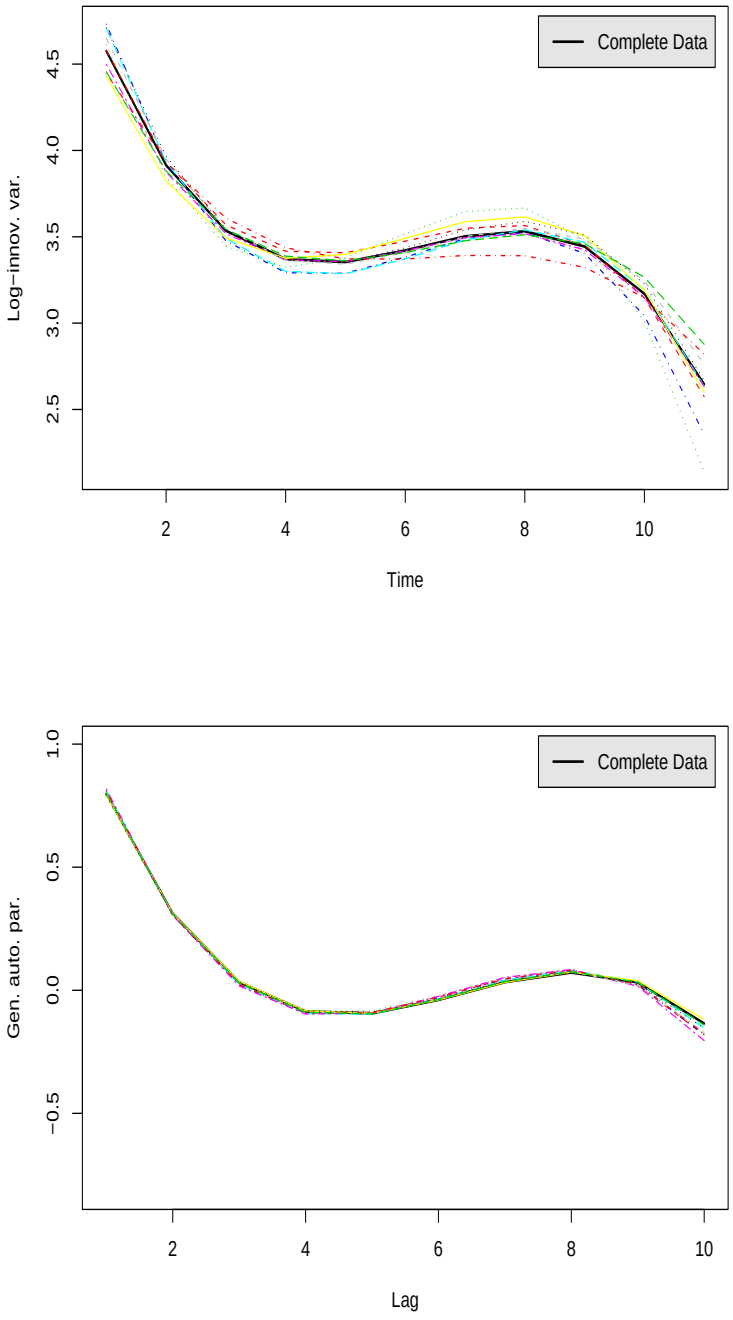

Figure 1: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 10% of the cattle data missing.

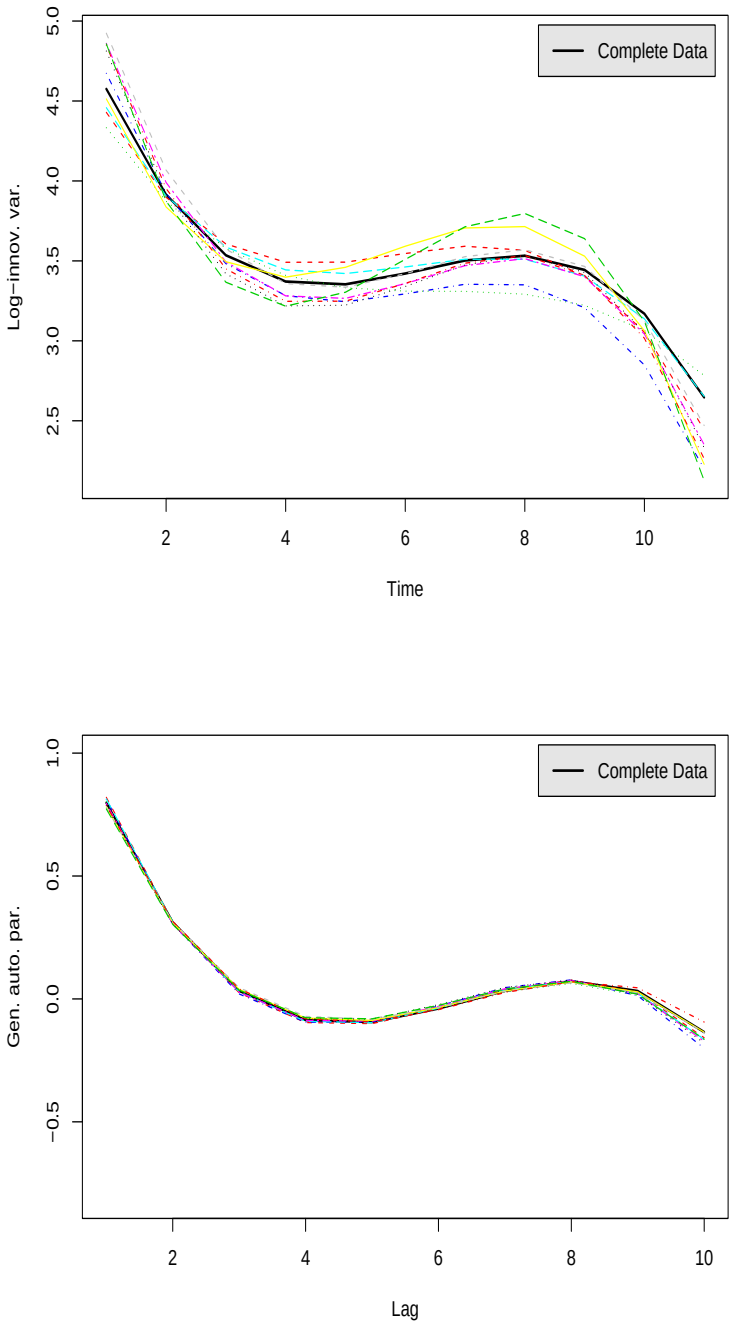

Figure 2: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 20% of the cattle data missing.

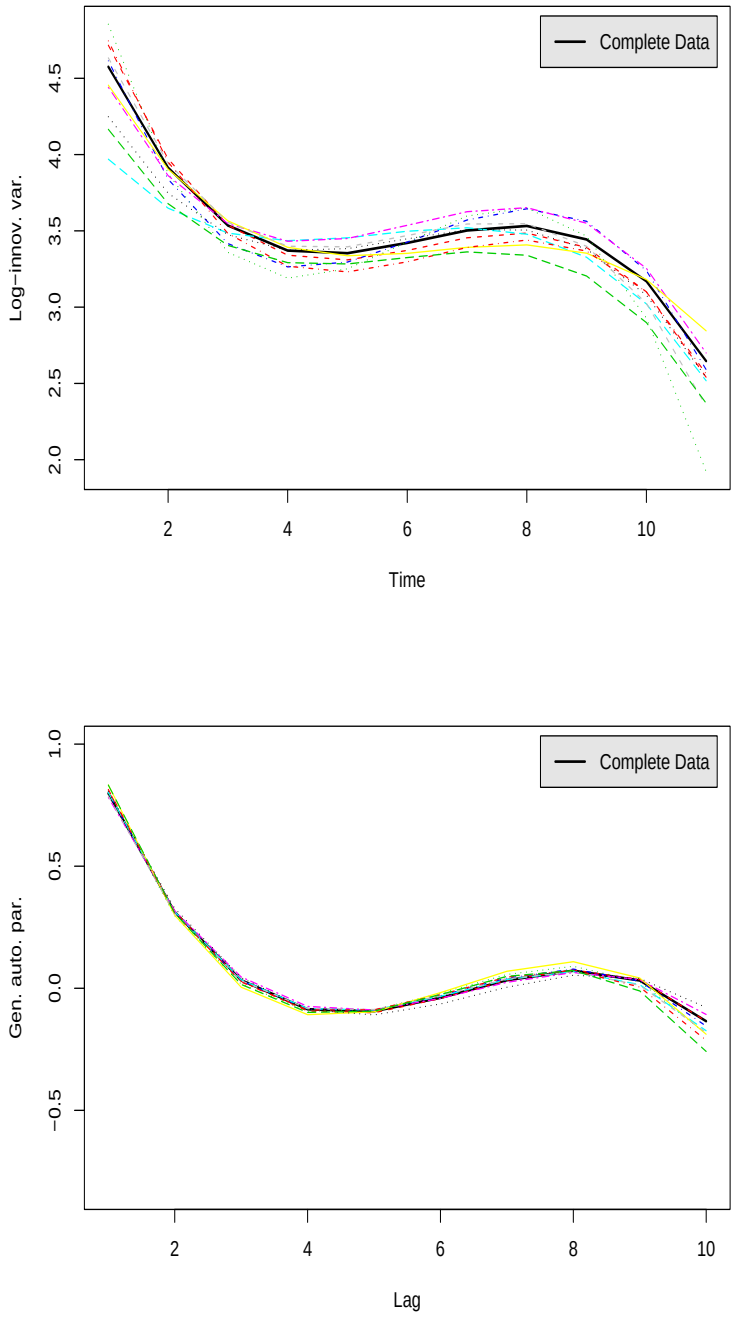

Figure 3: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 30% of the cattle data missing.

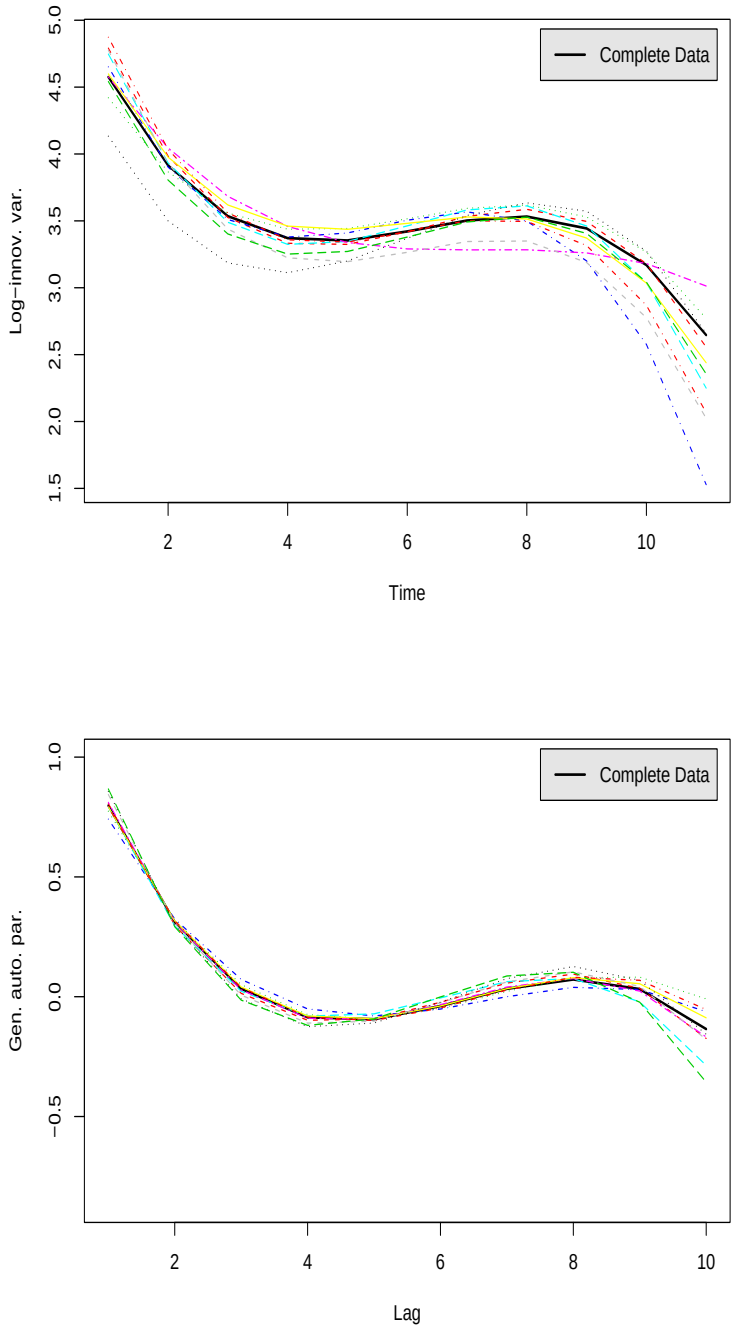

Figure 4: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 40% of the cattle data missing.

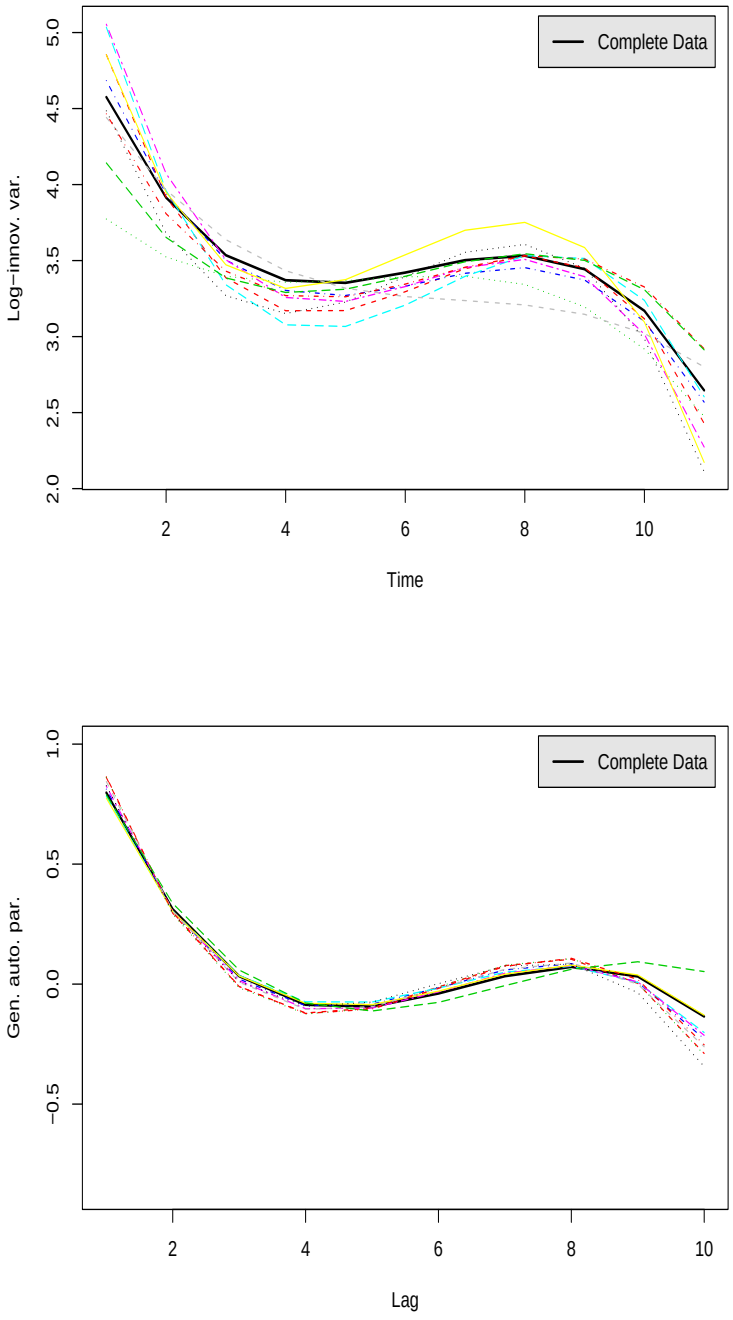

Figure 5: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 50% of the cattle data missing.

Table 1: Values of L_{max}, number of parameters and BIC for several models for Kenward's cattle data.

Model	L_{max}	Number of parameters	BIC
Unstructured Σ	-1019.69	66	75.35
Poly(3,3)	-1049.01	8	70.84
Poly(3,2)	-1080.08	7	72.80
Poly(3,1)	-1131.61	6	76.09
Poly(3,0)	-1215.35	5	81.59
Poly(3)	-1377.43	4	92.28
Unstructured AD(2)	-1035.98	30	72.47
Structured AD(2)	-1054.13	8	71.18
Stationary AR(2)	-1062.89	3	71.20
Structured AD(2) with $\lambda_1 = \lambda_2 = 1$	-1054.20	6	70.96

3.6 Simulation Studies

We conducted simulation studies mimicking the data pattern of Kenward's cattle data in the treatment group A. We considered the performance of the algorithm in the following scenarios: we generated the data which were equally spaced with $m_i = 11$, for all i, then we had 10%, 20%, 30%, 40%, or 50% of the simulation data missing completely at random. We adopted the saturated mean model and two cubic polynomials for the generalized autoregressive parameters and the logarithms of the innovation variances. We simulated with 400 replications from the same true model. That is

$$\log \tilde{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \epsilon_t,$$

and

$$\tilde{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \epsilon_{t,j},$$

where $t = 1, 2, \ldots, 11$ and $j = 1, 2, \ldots, t-1$.

The findings are shown in Table 2. The Monte Carlo standard deviations are in parentheses. The table shows good agreement between the ML estimators of (γ, λ)

Table 2: Simulation results for fixed-effects models: average maximum likelihood estimators and Monte Carlo standard deviations from the true model and five scenarios considered in the simulation study.

Model	Parameters	j=1	j=2	j=3	j=4
Ture model	λ_j	3.50	-1.18	0.22	-0.87
	γ_j	0.09	-0.54	0.46	-0.44
10% missing	λ_j	3.40(0.09)	-1.18(0.36)	0.13(0.37)	-0.87(0.34)
	γ_j	0.09(0.01)	-0.54(0.05)	0.45(0.06)	-0.44(0.09)
20% missing	λ_j	3.38(0.09)	-1.19(0.37)	0.11(0.37)	-0.86(0.39)
	γ_j	0.09(0.01)	-0.54(0.05)	0.45(0.07)	-0.45(0.09)
30% missing	λ_j	3.37(0.10)	-1.22(0.41)	0.12(0.42)	-0.86(0.41)
	γ_j	0.09(0.01)	-0.54(0.06)	0.45(0.08)	-0.45(0.11)
40% missing	λ_j	3.34(0.11)	-1.25(0.44)	0.10(0.49)	-0.86(0.44)
	γ_j	0.09(0.01)	-0.55(0.08)	0.45(0.10)	-0.46(0.13)
50% missing	λ_j	3.32(0.13)	-1.30(0.46)	0.06(0.52)	-0.86(0.52)
	γ_j	0.09(0.02)	-0.55(0.09)	0.45(0.11)	-0.46(0.14)

Values within parentheses are Monte Carlo standard deviations of 400 estimators

and the truth parameters in these five scenarios. The Monte Carlo standard deviation increases as the percentage of missing values increases.

Figure 6, Figure 7, Figure 8, Figure 9, and Figure 10 with 10 simulations for the above scenarios also show that the fitted log-innovation variances and generalized autoregressive parameters from unbalanced data using our generalized EM algorithm are close to the true values.

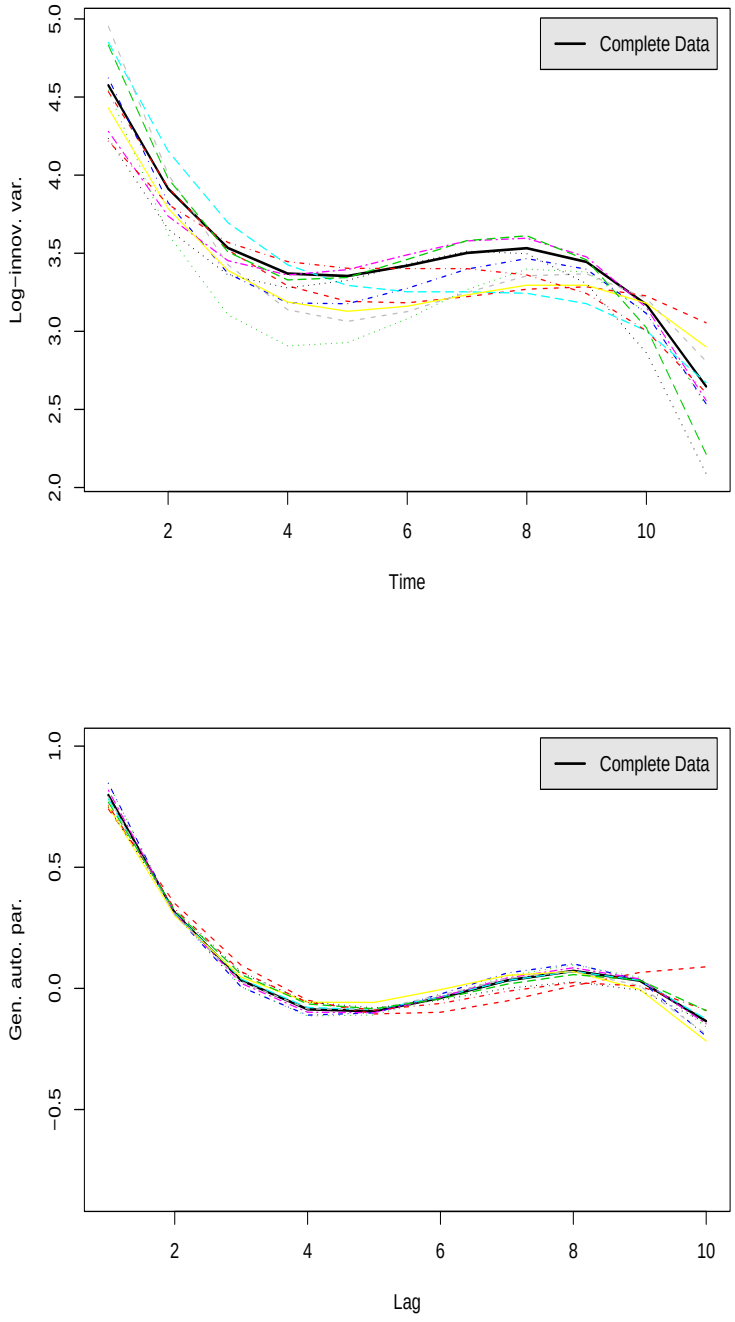

Figure 6: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 10% of the simulation data missing.

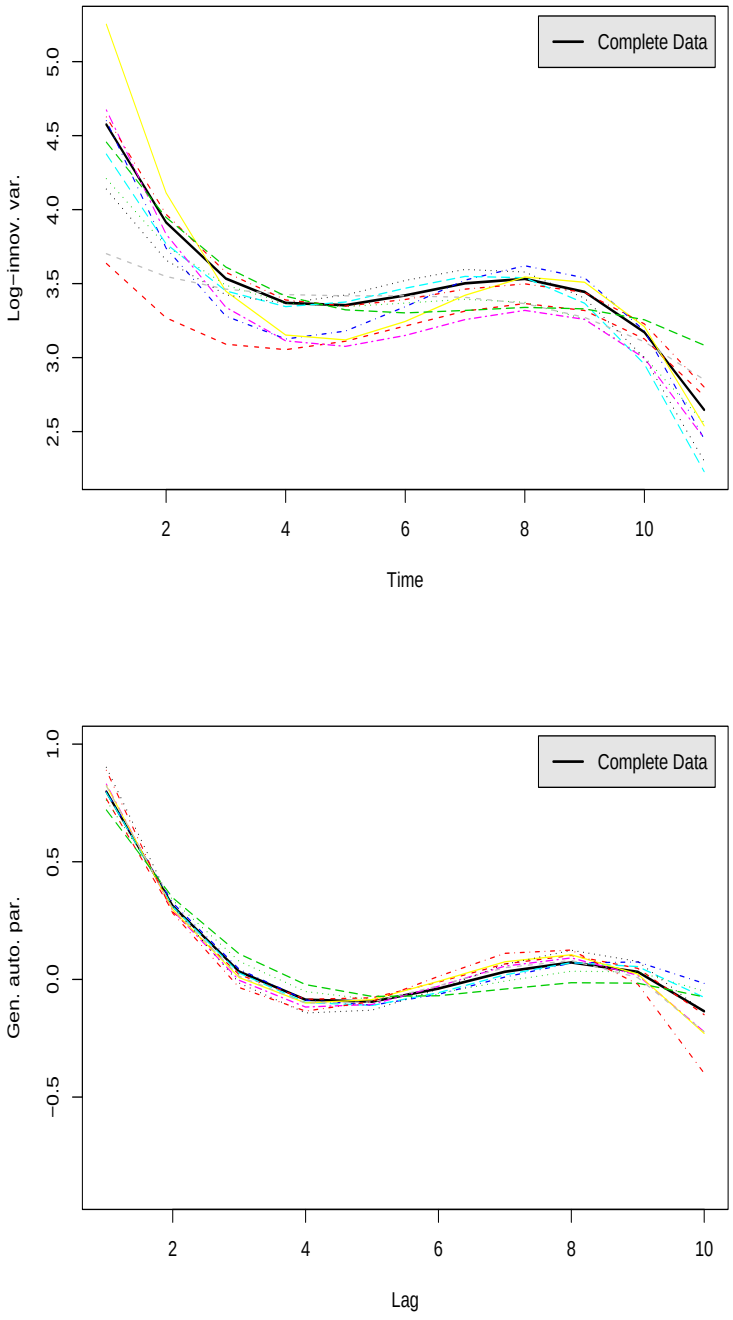

Figure 7: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 20% of the simulation data missing.

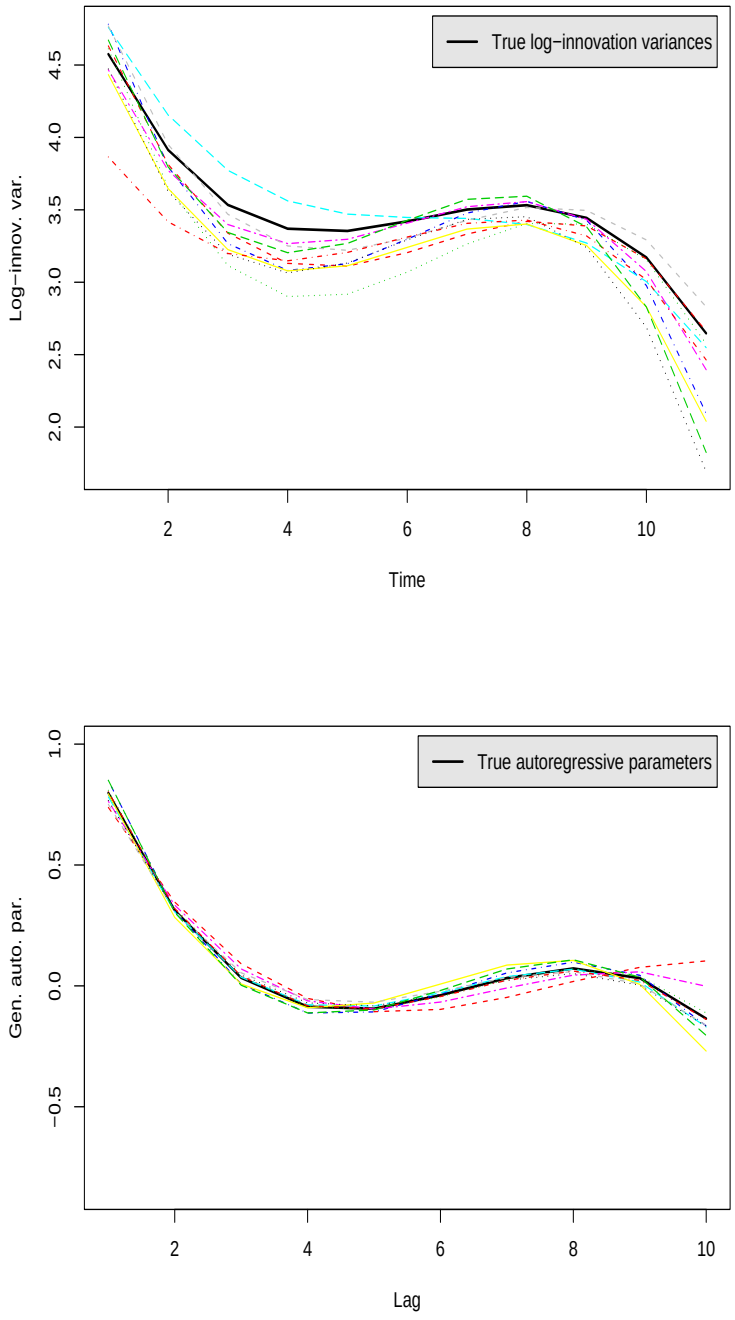

Figure 8: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 30% of the simulation data missing.

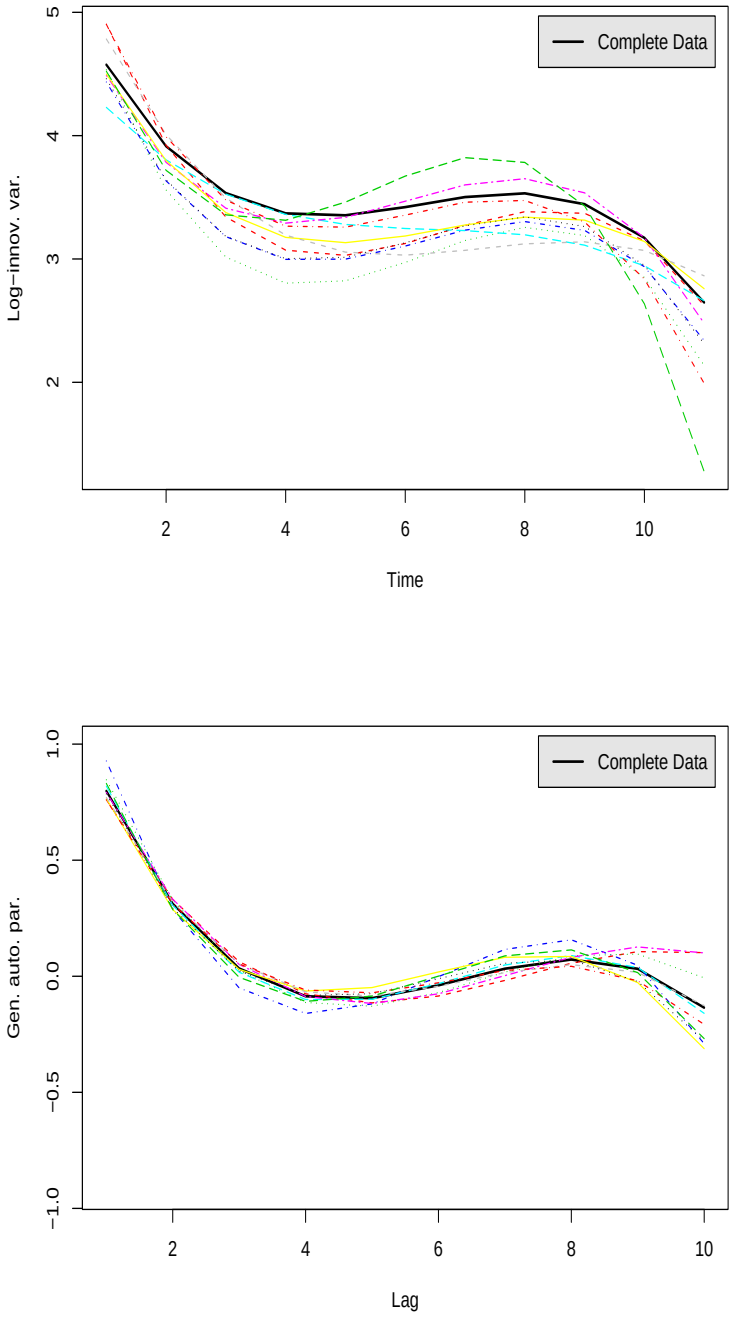

Figure 9: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 40% of the simulation data missing.

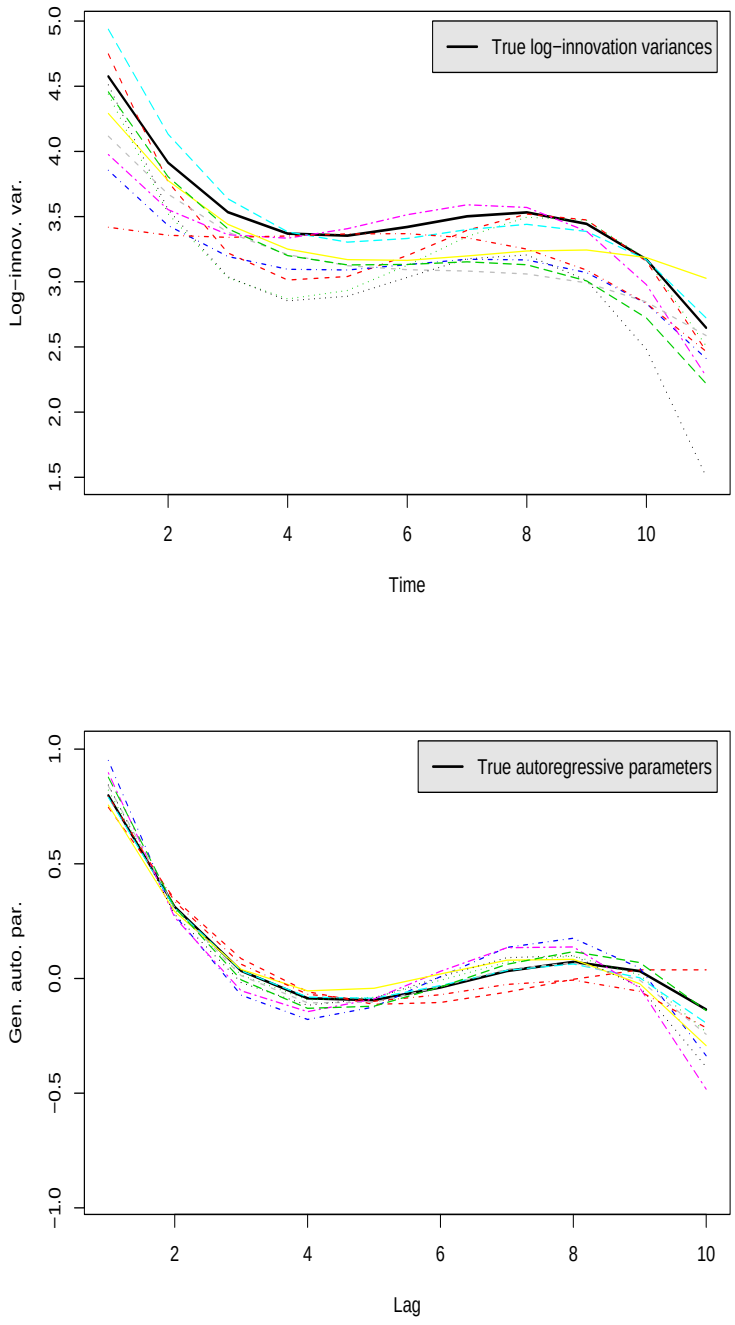

Figure 10: Fitted log-innovation variances and generalized autoregressive parameters for the fixed-effects model with 50% of the simulation data missing.

CHAPTER IV

MIXED-EFFECTS MODELS WITH UNBALANCED LONGITUDINAL DATA

4.1 Introduction

In this chapter, we reparameterize the within-subject covariance matrix in the mixed-effects model with unbalanced longitudinal data using the modified Cholesky decomposition and employ a generalized EM algorithm to compute the ML estimators. Section 4.2 presents the incomplete data mixed-effects model and details for the modified Cholesky decomposition of the covariance matrix. The algorithm to calculate the ML estimators is given in Section 4.3. Section 4.4 provides the asymptotic distribution of the ML estimators. Section 4.5 and 4.6 describe an application to Kenward's (1987) cattle data and simulation studies.

4.2 The Incomplete Data Model

Let y_i be a $m_i \times 1$ vector containing the responses for subject i, where $i = 1, \ldots, n$. The y_i are assumed to follow the model

$$y_i = X_i\beta + Z_i\alpha_i + e_i,$$

where X_i is a $m_i \times p$ known matrix, β is a $p \times 1$ vector of unknown regression parameters, Z_i is an $m_i \times b$ design matrix, the $b \times 1$ vector of between-subject random effects α_i are distributed as $N(0, B)$, and the $m_i \times 1$ vector of within-subject random effects e_i are independently distributed as $N(0, \Sigma_i)$. We assume b_i and e_i are independent. We further assume that e_i is a sub-vector of a $m \times 1$ vector e_i^* that corresponds to the same set of m observation times $t_1, \ldots, t_m$, for all i. It follows that Σ_i is a sub-matrix of $\Sigma_i^* = \text{var}(e_i^*)$. This model, called the incomplete data model, arises in situations

where a fixed number m of measurements, corresponding to different times, are to be collected on each of n subjects, but not all of the subjects' responses are observed (Jennrich and Schluchter, 1986).

The subject specific covariance matrix Σ_i^* can be modeled using covariates following the approach of Pourahmadi (1999, 2000). There exists a unique lower triangular matrix T_i with 1's as main diagonal entries and a unique diagonal matrix D_i with positive diagonal entries such that $T_i\Sigma_i^*T_i' = D_i$. The below-diagonal entries of T_i are the negatives of the autoregressive coefficients, ϕ_{itj}, in $\tilde{e}_{it}^* = \sum_{j=1}^{t-1}\phi_{itj}e_{ij}^*$, the linear least squares predictor of e_{ij}^* based on its predecessors $e_{i(t-1)}^*, \ldots, e_{i1}^*$. The diagonal entries of D_i are the innovation variances $\sigma_{it}^2 = \text{var}(e_{it}^* - \tilde{e}_{it}^*)$, where $1 \le t \le m$ and $1 \le i \le n$. The parameters ϕ_{itj} and $\log\sigma_{it}^2$ are unconstrained and are modeled as $\phi_{itj} = z_{itj}'\gamma$ and $\log(\sigma_{it}^2) = u_{it}'\lambda$, $1 \le j \le t-1$, $1 \le t \le m$, where z_{itj} and u_{it} are covariates for covariance matrices and $\gamma \in R^q$ and $\lambda \in R^r$ are corresponding regression parameters of interest. We assume that there is no missing value in these covariates, which is the case if they only depend on baseline covariates and scheduled observation times.

The between-subject covariance matrix is denoted by the general parameter vector δ, i.e., $B = B(\delta)$.

The model is flexible. The $\log(\sigma_{it}^2)$ and ϕ_{itj} can have nonlinear functions of their parameters γ, λ respectively. For example, the matrices T with entries $\phi_{i,j}$ given respectively by

$$\gamma^{i-j}, \gamma^{(t_i - t_{i-j})^\theta}, \gamma_j^{f(t_i,\lambda_j)-f(t_{i-j},\lambda_j)}, \gamma_{i-j}$$

are correspond to $AR(1)$, damped exponential (Diggle, 1988; Munoz et al., 1992), structured antedependence (Zimmerman and Nunez-Anton, 1997) and banded (Pourahmadi, 1999).

4.3 Algorithm for Maximum Likelihood

This section derives a generalized EM algorithm (Dempster et al., 1977; Jennrich and Schluchter, 1986; Laird et al., 1987) to compute the ML estimators. The iterative algorithm consists of three parts. The first part uses generalized least squares to update fixed effects β:

$$\tilde{\beta} = \Big(\sum_{i=1}^{n} X_i'\Sigma_i^{-1}X_i\Big)^{-1}\Big\{\sum_{i=1}^{n} X_i'\Sigma_i^{-1}(y_i - Z_i\alpha_i)\Big\}, \tag{4.1}$$

which is equivalent to maximize the likelihood with respect to β while holding B, γ and λ fixed. The second part estimates the between-subject covariance matrix B. The third part comprises one iteration of a generalized EM algorithm to update λ and γ. Both the second and third parts use (e_i^*, α_i) as complete data and $\eta_i = y_i - X_i\beta = Z_i\alpha_i + e_i$ as observed data, and assuming β is equal to its current value. The updates of γ and λ are obtained separately using weighted least squares and one step of Newton-Raphson.

Now we give details of the second part and third part of the algorithm. Minus twice the log likelihood function for complete data, except for a constant, is given by

$$\begin{aligned} -2l &= \sum_{i=1}^{n}(\log|\Sigma_i^*| + e_i^{*'}\Sigma_i^{*-1}e_i^* + \log|B| + \alpha_i'B^{-1}\alpha_i) \\ &= \sum_{i=1}^{n}\{[\log|\Sigma_i^*| + \mathrm{tr}(\Sigma_i^{*-1}V_i) + \log|B| + \mathrm{tr}(B^{-1}U_i)\}. \end{aligned}$$

where $V_i = e_i^*e_i^{*'}$ and $U_i = \alpha_i\alpha_i'$. Let Q be the expected log likelihood given the observed data and the current parameter values. Then

$$-2Q = \sum_{i=1}^{n}\{\log|\Sigma_i^*| + \mathrm{tr}(\Sigma_i^{*-1}\widehat{V}_i) + \log|B| + \mathrm{tr}(B^{-1}\widehat{U}_i)\},$$

where $\widehat{V}_i = E(e_i^*e_i^{*'}|\eta_i) = \widehat{e_i^*}\widehat{e_i^*}' + \mathrm{var}(e_i^*|\eta_i)$ with $\widehat{e_i^*} = E(e_i^*|\eta_i)$, and $\widehat{U}_i = E(\alpha_i\alpha_i'|\eta_i) = \widehat{\alpha_i}\widehat{\alpha_i}' + \mathrm{var}(\alpha_i|\eta_i)$ with $\widehat{\alpha_i} = E(\alpha_i|\eta_i)$. The calculation of $\widehat{\alpha_i}$ and $\widehat{U}_i$ is as follows:

$$E(\alpha_i|\eta_i) = BZ_i'(\Sigma_i + Z_iBZ_i')^{-1}\eta_i. \tag{4.2}$$

and

$$\begin{aligned} \mathrm{var}(\alpha_i|\eta_i) &= B - BZ_i'(\Sigma_i + Z_iBZ_i')^{-1}Z_iB \\ &= (Z_i'\Sigma_i^{-1}Z_i + B^{-1})^{-1}. \end{aligned} \tag{4.3}$$

The second equality in (4.3) is based on a standard matrix calculation, see for example, p358 of Pawitan (2001). The calculation of $\widehat{e_i^*}$ and $\widehat{V_i}$ follows standard results for multivariate normal distributions. For example, if

$$e_i^* = \begin{pmatrix} e_i \\ e_i^+ \end{pmatrix} \sim N(0, \Sigma_i^*), \qquad \Sigma_i^* = \begin{pmatrix} \Sigma_{i11}^* & \Sigma_{i12}^* \\ \Sigma_{i21}^* & \Sigma_{i22}^* \end{pmatrix}, \qquad \Sigma_{i11}^* = \Sigma_i,$$

then

$$\begin{pmatrix} e_i \\ e_i^+ \\ \eta_i \end{pmatrix} \sim N(0, \begin{pmatrix} \Sigma_{i11}^* & \Sigma_{i12}^* & \Sigma_{i11}^* \\ \Sigma_{i21}^* & \Sigma_{i22}^* & \Sigma_{i21}^* \\ \Sigma_{i11}^* & \Sigma_{i12}^* & Z_iBZ_i' + \Sigma_{i11}^* \end{pmatrix}).$$

Hence,

$$E(e_i^*|\eta_i) = \begin{pmatrix} \Sigma_{i11}^* \\ \Sigma_{i21}^* \end{pmatrix} (\Sigma_{i11}^* + Z_iBZ_i')^{-1}\eta_i \tag{4.4}$$

and

$$\mathrm{var}(e_i^*|\eta_i) = \Sigma_i^* - \begin{pmatrix} \Sigma_{i11}^* \\ \Sigma_{i21}^* \end{pmatrix} (\Sigma_{i11}^* + Z_iBZ_i')^{-1} \begin{pmatrix} \Sigma_{i11}^* \\ \Sigma_{i21}^* \end{pmatrix}'. \tag{4.5}$$

In the EM algorithm we need to minimize $-2Q$ with respect to γ and λ. We now derive two expressions of $-2Q$ that are useful in solving the optimization problem. First note that $|\Sigma_i^*| = |D_i| = \prod_{t=1}^m \sigma_{it}^2$. Since $\Sigma_i^{*-1} = T_i'D_i^{-1}T_i$ and the t-th row of $T_ie_i^*$ is

$$e_{it}^* - \sum_{j=1}^{t-1} \phi_{itj}e_{ij}^* = e_{it}^* - \sum_{j=1}^{t-1} e_{ij}^* z_{itj}'\gamma,$$

we have that

$$e_i^{*\prime}\Sigma_i^{*-1}e_i^* = e_i^{*\prime}T_i'D_i^{-1}T_ie_i^* = \sum_{t=1}^m \sigma_{it}^{-2}\mathrm{RS}_{it},$$

where the residual squared $\mathrm{RS}_{it} = (e^*_{it} - \sum_{j=1}^{t-1} e^*_{ij} z'_{itj}\gamma)^2$. Denote $Z'_{it} = (z_{it1}, \ldots, z_{it(t-1)})$ and recall that $V_i = (V_{ijj'}) = e^*_i e^{*'}_i$. Let $V_{it} = V_i[\cdot, t]$, $V^{(t-1)}_{it} = V_i[1{:}(t-1), t]$ and $V^{(t-1)}_i = V_i[1{:}(t-1), 1{:}(t-1)]$ be appropriate sub-matrices of V_i. Then

$$\begin{aligned}\mathrm{RS}_{it} &= e^{*2}_{it} - 2\sum_{j=1}^{t-1} \gamma' z_{itj} e^*_{ij} e^*_{it} + \gamma'\Big(\sum_{j,j'=1}^{t-1} z_{itj} e^*_{ij} e^*_{ij'} z'_{itj'}\Big)\gamma \\ &= V_{itt} - 2\gamma' Z'_{it} V^{(t-1)}_{it} + \gamma' Z'_{it} V^{(t-1)}_i Z_{it}\gamma.\end{aligned}$$

Note that if $t = 1$ we have $V^{(t-1)}_{it} = 0$ and $V^{(t-1)}_i = 0$. The above discussion yields two expressions of the log likelihood of complete data

$$\begin{aligned}-2l &= \sum_{i=1}^{n}\sum_{t=1}^{m}\Big(\log\sigma^2_{it} + \frac{\mathrm{RS}_{it}}{\sigma^2_{it}}\Big) + \sum_{i=1}^{n}\{\log|B| + \mathrm{tr}(B^{-1}U_i)\} \\ &= \sum_{i=1}^{n}\sum_{t=1}^{m}\Big(\log\sigma^2_{it} + \frac{V_{itt}}{\sigma^2_{it}}\Big) \\ &\quad + \sum_{i=1}^{n}\sum_{t=1}^{m}\sigma^{-2}_{it}(-2\gamma' Z'_{it} V^{(t-1)}_{it} + \gamma' Z'_{it} V^{(t-1)}_i Z_{it}\gamma) + \sum_{i=1}^{n}\{\log|B| + \mathrm{tr}(B^{-1}U_i)\}.\end{aligned}$$

It follows that the expected log likelihood given observed data has two expressions

$$\begin{aligned}-2Q &= \sum_{i=1}^{n}\sum_{t=1}^{m}\Big(\log\sigma^2_{it} + \frac{\widehat{\mathrm{RS}}_{it}}{\sigma^2_{it}}\Big) + \sum_{i=1}^{n}\{\log|B| + \mathrm{tr}(B^{-1}\widehat{U}_i)\} \\ &= \sum_{i=1}^{n}\sum_{t=1}^{m}\Big(\log\sigma^2_{it} + \frac{\widehat{V}_{itt}}{\sigma^2_{it}}\Big) \\ &\quad + \sum_{i=1}^{n}\sum_{t=1}^{m}\sigma^{-2}_{it}(-2\gamma' Z'_{it}\widehat{V}^{(t-1)}_{it} + \gamma' Z'_{it}\widehat{V}^{(t-1)}_i Z_{it}\gamma) + \sum_{i=1}^{n}\{\log|B| + \mathrm{tr}(B^{-1}\widehat{U}_i)\},\end{aligned}$$

where $\widehat{\mathrm{RS}}_{it} = E(\mathrm{RS}_{it}|e_i)$ and $\widehat{V}_i$ with sub- and super-scripts are appropriate sub-matrices of $\widehat{V}_i$. We can also show that $\widehat{\mathrm{RS}}_{it}$ is the (t,t)-th element of the matrix $T_i\widehat{V}_i T'_i$.

For updating B with fixed γ and λ, $-2Q$ is minimized by

$$\tilde{B} = \frac{1}{n}\sum_{i=1}^{n}\widehat{U}_i. \tag{4.6}$$

The update of γ and λ proceeds as follows. For fixed λ, $-2Q$ is a quadratic form in γ and is minimized by

$$\tilde{\gamma} = \Big(\sum_{i=1}^{n}\sum_{t=1}^{m} \sigma_{it}^{-2} Z_{it}' \widehat{V}_i^{(t-1)} Z_{it}\Big)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{m} \sigma_{it}^{-2} Z_{it}' \widehat{V}_{it}^{(t-1)}. \tag{4.7}$$

For fixed γ, optimization of $-2Q$ over λ does not have a closed-form expression and we rely on the Newton-Raphson algorithm. Since $\log \sigma_{it}^2 = u_{it}'\lambda$, simple calculation yields

$$\frac{\partial Q}{\partial \lambda} = -\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m}\Big(1 - \frac{\widehat{\mathrm{RS}}_{it}}{\sigma_{it}^2}\Big)u_{it}$$

and

$$\frac{\partial^2 Q}{\partial \lambda \partial \lambda'} = -\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{\widehat{\mathrm{RS}}_{it}}{\sigma_{it}^2}u_{it}u_{it}'.$$

The Newton-Raphson algorithm updates the current values $\lambda^{(0)}$ to $\lambda^{(1)}$ using

$$\lambda^{(1)} = \lambda^{(0)} + \Delta\lambda, \qquad \Delta\lambda = -\Big(\frac{\partial^2 Q}{\partial \lambda \partial \lambda'}\Big)^{-1}\frac{\partial Q}{\partial \lambda}. \tag{4.8}$$

For the generalized EM algorithm, we only do one step instead of full iteration of the Newton-Raphson. We need make sure that the log likelihood increases at each step of the generalized EM algorithm, using partial stepping such as step-halving if necessary. Step-halving works as follows. If $Q(\lambda^{(1)}) \leq Q(\lambda^{(0)})$, we replace $\Delta\lambda$ by its half in the update $\lambda^{(1)} = \lambda^{(0)} + \Delta\lambda$, and continue doing so until $Q(\lambda^{(1)}) > Q(\lambda^{(0)})$.

The steps of the algorithm are summarized as follows:

(i) Initialization: set $\alpha = 0$, $b \times 1$ vector,$B = I$ and $\Sigma_i^* = I$, $i = 1, \ldots, n$.

(ii) Using the current estimates of α, B, γ and λ (or Σ_i^* in the first iteration), compute the updated estimate $\tilde{\beta}$ of β using equation (4.1).

(iii) Using the current estimates of β, B, γ and λ,update α using equation (4.2).

(iv) Compute $\widehat{U}_i$, $i = 1, \ldots, n$, where the relevant conditional expectations are calculated using (4.2) and (4.3).

(v) Compute $\widehat{V}_i$, $i = 1, \ldots, n$, where the relevant conditional expectations are calculated using (4.4) and (4.5).

(vi) Using the current estimate of α, β, γ and λ, update the current estimate $B^{(0)}$ to $B^{(1)}$ using (4.6).

(vii) Using the current estimates of α, β, B and λ, update γ using (4.7).

(viii) Using the current estimates of α, β, B and γ, update the current estimate $\lambda^{(0)}$ to $\lambda^{(1)}$ using one step of Newton-Raphson as (4.8). Use step-halving to guarantee that the likelihood is increased.

(ix) Iterate (ii)–(viii) until convergence.

4.4 Observed Information Matrix

Since we are in a parametric framework, the standard likelihood theory applies. The ML estimators of $\beta, \gamma, \lambda, \delta$ are consistent and normally distributed.

One drawback of the EM algorithm is that it does not automatically provide standard errors for the estimates. If the observed data likelihood $L(\theta; y)$ is available, $\hat{\theta}$ can be obtained, then we can find the standard errors from the observed Fisher information $I(\hat{\theta}; y)$. The asymptotic variance-covariance matrix of the maximum likelihood estimators $\hat{\theta}$ can be computed using the inverse of the observed information matrix $I(\hat{\theta}; y)$, evaluated at the $\hat{\theta}$.

We follow Oakes (1999) and Pawitan (2001) as in Section 3.4. The observed score in terms of the first derivative of Q can be express as

$$S(\theta^0; y) = \frac{\partial Q(\theta|\theta^0)}{\partial \theta}|_{\theta=\theta^0}.$$

This relationship holds at each y_i for independent data $y_1, \ldots, y_n$; therefore

$$S(\theta^0; y_i) = \frac{\partial Q_i(\theta|\theta^0)}{\partial \theta}|_{\theta=\theta^0},$$

where $Q_i = E\{\log L(\theta; y_i^*)|y_i, \theta^0\}$. The estimate of the observed Fisher information is

$$I(\tilde{\theta}; y) \approx \sum_i S(\tilde{\theta}; y_i) S^t(\tilde{\theta}; y_i).$$

We can estimate the variance-covariance matrix of $\tilde{\theta} = (\tilde{\gamma}, \tilde{\lambda}, \tilde{\delta})$, since the observed Fisher information

$$\begin{aligned} I(\tilde{\theta}; y) &\approx \sum_i S(\tilde{\theta}; y_i) S^t(\tilde{\theta}; y_i) \\ &\approx \begin{pmatrix} \sum_i S(\tilde{\gamma}; y_i) S^t(\tilde{\gamma}; y_i) & \sum_i S(\tilde{\gamma}; y_i) S^t(\tilde{\lambda}; y_i) & \sum_i S(\tilde{\gamma}; y_i) S^t(\tilde{\delta}; y_i) \\ \sum_i S(\tilde{\lambda}; y_i) S^t(\tilde{\gamma}; y_i) & \sum_i S(\tilde{\lambda}; y_i) S^t(\tilde{\lambda}; y_i) & \sum_i S(\tilde{\lambda}; y_i) S^t(\tilde{\delta}; y_i) \\ \sum_i S(\tilde{\delta}; y_i) S^t(\tilde{\gamma}; y_i) & \sum_i S(\tilde{\delta}; y_i) S^t(\tilde{\lambda}; y_i) & \sum_i S(\tilde{\delta}; y_i) S^t(\tilde{\delta}; y_i) \end{pmatrix}, \end{aligned} \tag{4.9}$$

where

$$S(\tilde{\theta}; y_i) = [\frac{\partial Q_i(\theta|\theta^{(0)})}{\partial \theta}]_{\theta=\tilde{\theta}}.$$

We have that

$$S(\tilde{\gamma}; y_i) = [\frac{\partial Q_i}{\partial \gamma}]_{\gamma=\tilde{\gamma}} = [\sum_{t=1}^{m} \sigma_{it}^{-2} (Z_{it}' \widehat{V}_{it}^{(t-1)} - Z_{it}' \widehat{V}_i^{(t-1)} Z_{it}\gamma)]|_{\gamma=\tilde{\gamma}}$$

and

$$S(\tilde{\lambda}; y_i) = [\frac{\partial Q_i}{\partial \lambda}]_{\lambda=\tilde{\lambda}} = [-\frac{1}{2} \sum_{t=1}^{m} \Big(1 - \frac{\widehat{\mathrm{RS}}_{it}}{\sigma_{it}^2}\Big) u_{it}]|_{\lambda=\tilde{\lambda}}.$$

The k-th component of the $S(\tilde{\delta}; y_i)$ is

$$S_k(\tilde{\delta}; y_i) = [\frac{\partial Q_i}{\partial \delta_k}]_{\delta=\tilde{\delta}} = [-\frac{1}{2} tr\{(B^{-1} - B^{-1} \widehat{U}_i B^{-1}) \dot{B}_k\}]|_{\delta=\tilde{\delta}}$$

where $\dot{B}_k = \partial B / \partial \delta_k$. Then the asymptotic variances of the ML estimator $\tilde{\theta}$ can be calculated using the inverse of the observed Fisher information (4.9), evaluated at the ML estimator $\tilde{\theta}$.

Since $\tilde{\beta}$ and $(\tilde{\gamma}, \tilde{\lambda}, \tilde{\delta})$ are asymptotically independent and consistent estimators for $(\beta', \gamma', \lambda', \delta')'$, the asymptotic covariance matrix of $\tilde{\beta}$ can be estimated by $(\sum_i X_i' \tilde{\Sigma}^{-1} X_i)^{-1}$.

4.5 An Application to Kenward's Cattle Data

Kenward (1987) showed an experiment to study the effect of treatments on intestinal parasites of cattle. Thirty cattle were randomly assigned to each of the treatment groups A and B. The weights of the cattle were measured 11 times over a 133-day period. The intervals between consecutive measures were two weeks, except the last one was one week. The measures were taken at the same set of times and there were no missing observations. The times were rescaled as $t = 1, 2, \ldots, 11$. This is a balanced longitudinal data. Pourahmadi (1999) used a saturated mean model and reparameterized the covariance matrix with the modified Cholesky decomposition for the treatment group A. He suggested that both the generalized autoregressive parameters and the logarithms of the innovation variances were cubic functions of the lag or time. That is,

$$\tilde{\phi}_{tj} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \epsilon_{tj},$$

and

$$\log \tilde{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \epsilon_t,$$

where $t = 1, 2, \ldots, 11$ and $j = 1, 2, \ldots, t-1$. Thus, he did not explicitly take the between-cattle variation into account.

Pan and MacKenzie (2007) compared Pourahmadi's fixed-effects model with several linear mixed-effects models. They adopted the cubic polynomial regression scheme for the autoregressive coefficients and the innovation variances in the Cholesky decomposition of the within-subject covariance matrix for group A. They assumed that the random effects $\alpha_i \sim N(0, B(\delta))$, where $B(\delta) = \text{diag}(\delta_1^2, \delta_2^2, \ldots, \delta_{\text{b}}^2)$. Both the BIC and AIC criteria select the random intercept linear mixed-effects model ($b = 1$) as the best model. A formal test shows there is between-cattle variation which

should be taken into account. Furthermore, they compared their random intercept linear mixed-effects model ($b = 1$) with several menu-selection models including independent, compound symmetry, $AR(1)$, $AR(2)$ or $ARMA(1,1)$ as the within-subject covariance matrix. The BIC and AIC show that their model is much better than the menu-selection methods considered.

The objective of our analysis is to apply the modified Cholesky decomposition to unbalanced longitudinal data. We had 10%, 20%, 30%, 40%, or 50% of the cattle data missing completely at random. Then we adopted Pan and MacKenzie's scheme using the random intercept linear mixed-effects model ($b = 1$) and estimated the ML estimators with the generalized EM algorithm. We repeated this procedure ten times. The results are shown in Figure 11, Figure 12, Figure 13, Figure 14, and Figure 15. The fitted log-innovation variances and generalized autoregressive parameters from unbalanced data using our generalized EM algorithm are close to those from the complete cattle data using Pourahmadi (1999)'s algorithm.

4.6 Simulation Studies

We conducted simulation studies mimicking the data pattern of Kenward's cattle data in the treatment group A. We considered the performance of the algorithm in the following scenarios: we generated the data which were equally spaced with $m_i = 11$, for all i, then we had 10%, 20%, 30%, 40%, or 50% of the simulation data missing completely at random. We adopted saturated mean for the random intercept linear mixed-effects model ($b = 1$). We used two cubic polynomials for the generalized autoregressive parameters and the logarithms of the innovation variances. We simulated with 400 replications from the same true model.

The findings are shown in Table 3. The Monte Carlo standard deviations are in parentheses. The table shows good agreement between the ML estimators of (γ, λ)

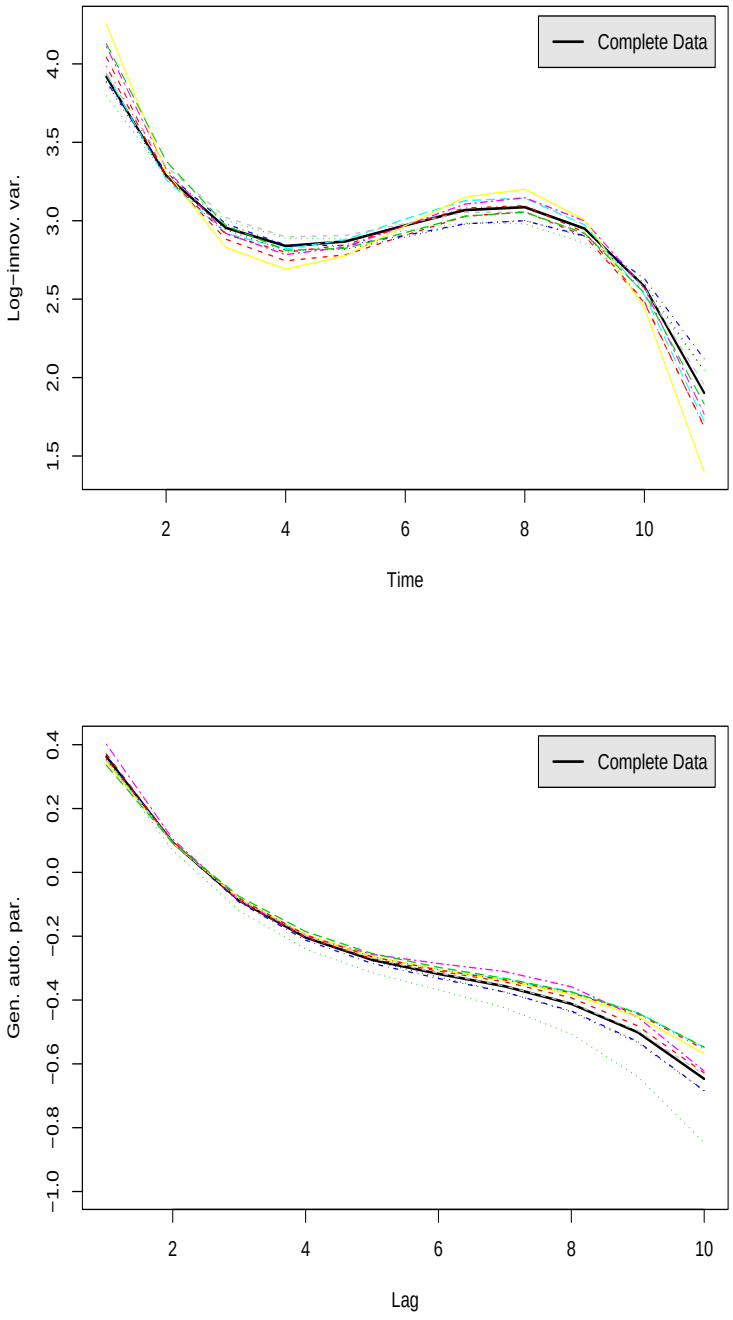

Figure 11: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 10% of the cattle data missing.

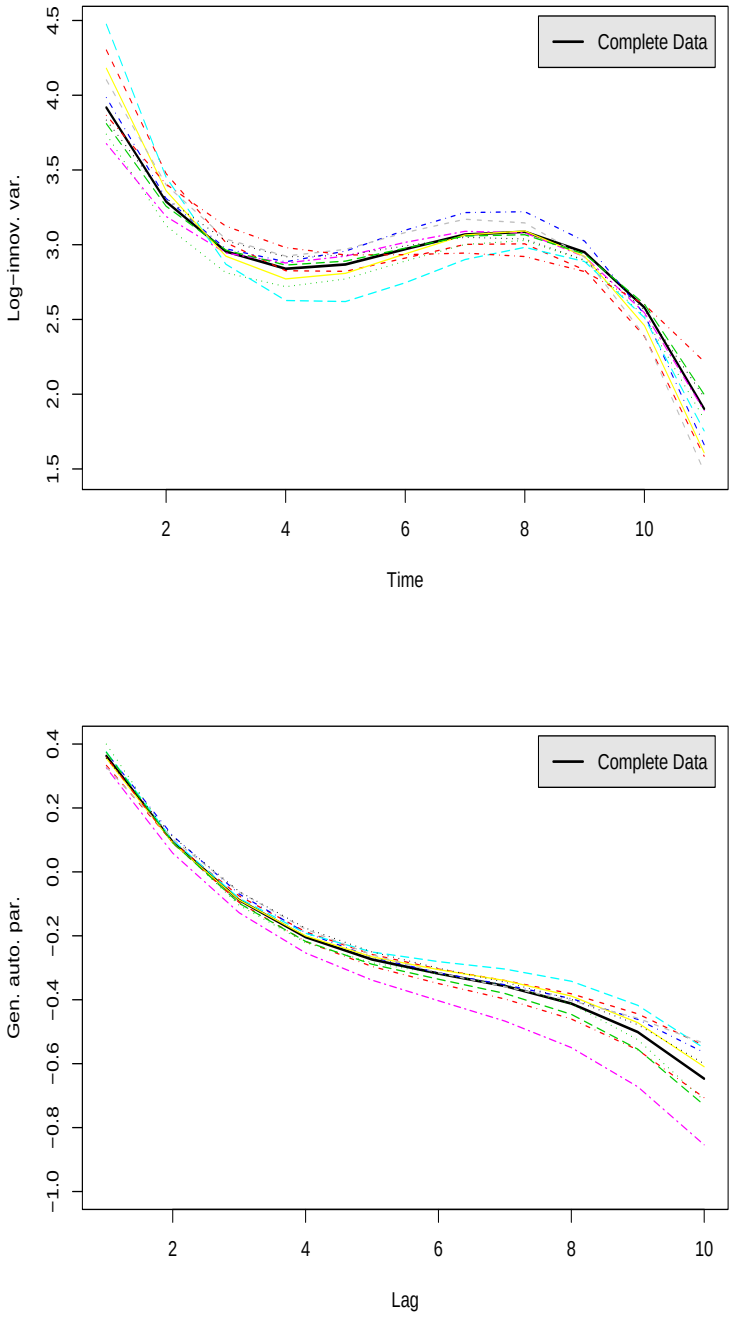

Figure 12: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 20% of the cattle data missing.

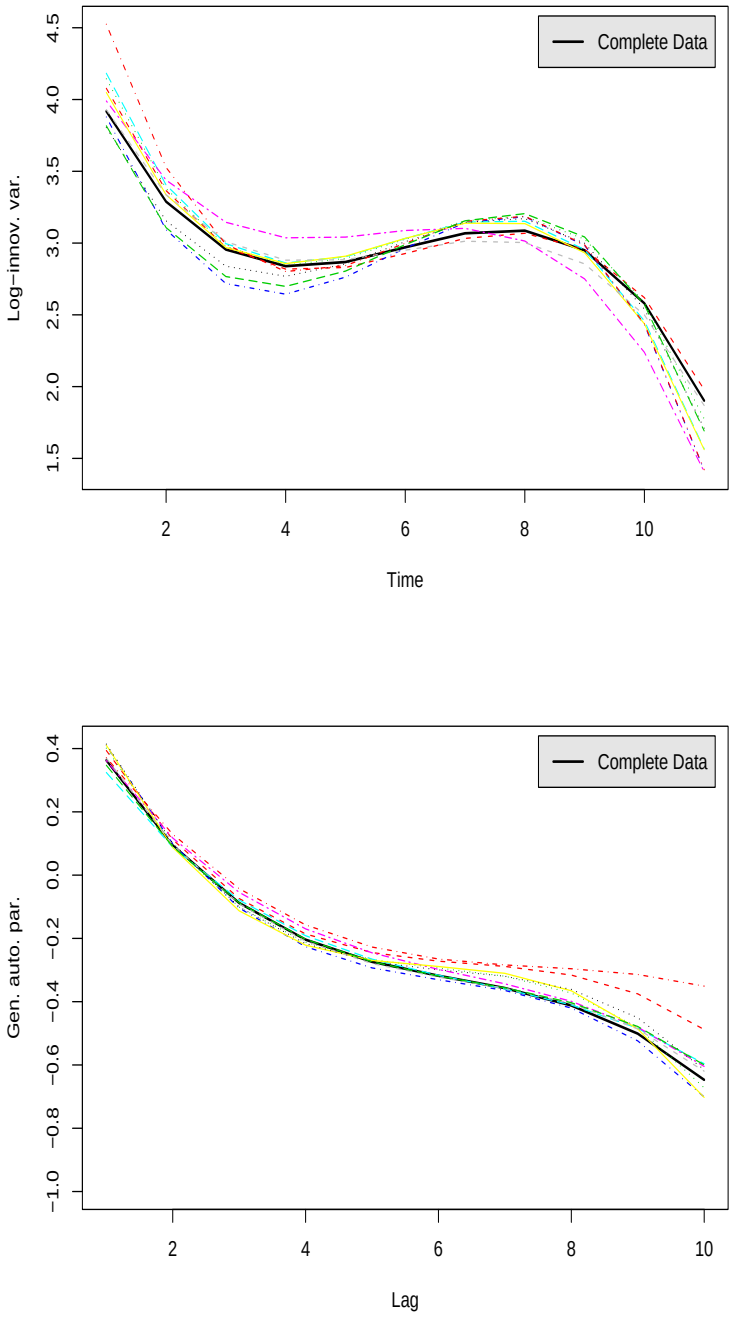

Figure 13: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 30% of the cattle data missing.

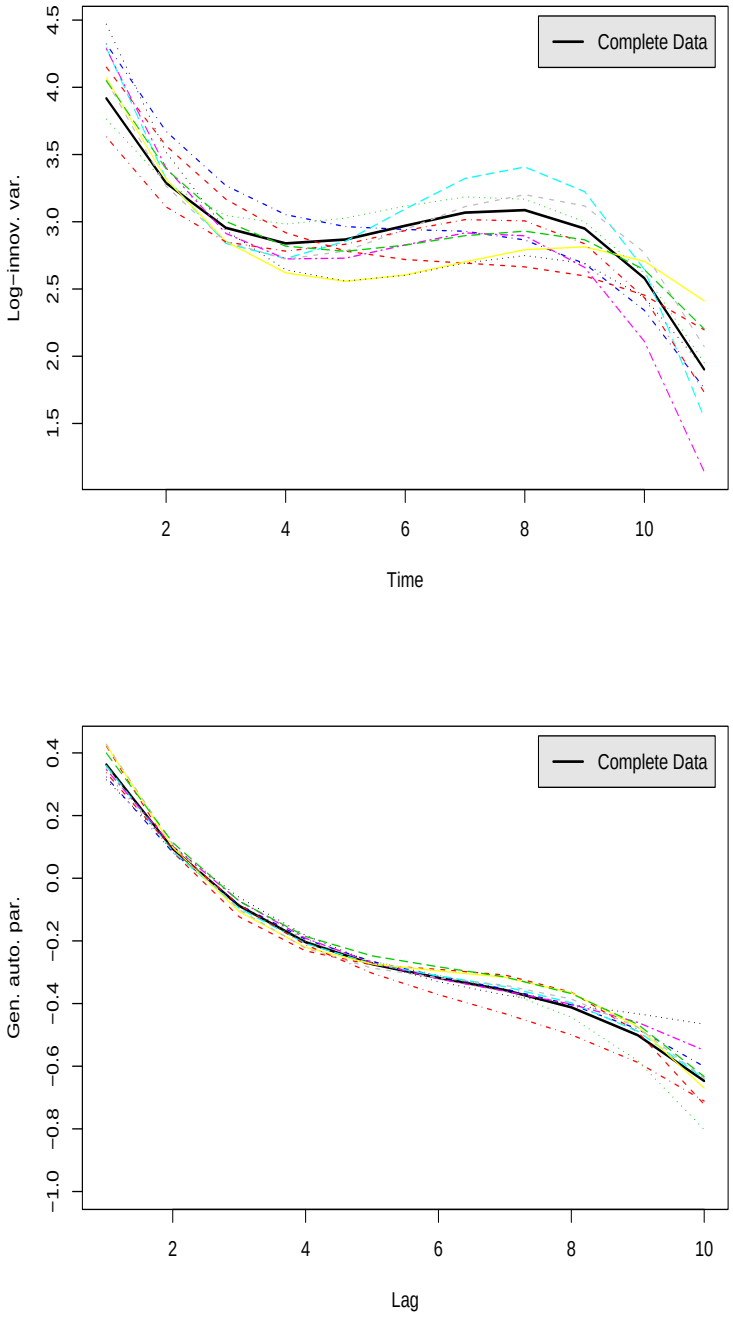

Figure 14: Fitted log-innovation variances and generalized autoregressive parameters fro the mixed-effects model with 40% of the cattle data missing.

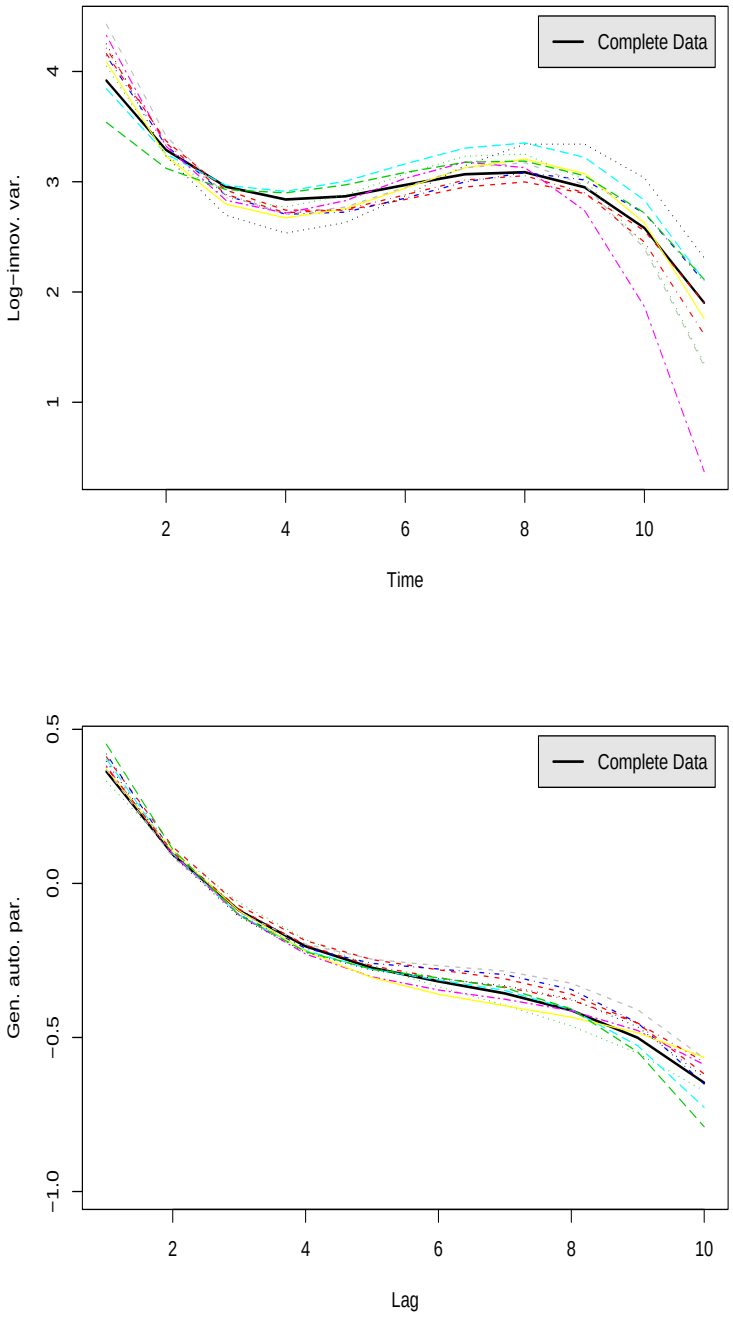

Figure 15: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 50% of the cattle data missing.

Table 3: Simulation results for mixed-effects models: average maximum likelihood estimators and Monte Carlo standard deviations from the true model and five scenarios considered in the simulation study.

Model	Parameters	j=1	j=2	j=3	j=4
Ture model	λ_j	2.95	-1.17	-0.07	-0.99
	γ_j	-0.23	-0.85	0.18	-0.19
10% missing	λ_j	2.90(0.11)	-1.22(0.38)	-0.06(0.40)	-0.98(0.37)
	γ_j	-0.24(0.05)	-0.85(0.10)	0.17(0.07)	-0.19(0.08)
20% missing	λ_j	2.90(0.12)	-1.23(0.39)	-0.08(0.41)	-0.98(0.40)
	γ_j	-0.23(0.06)	-0.84(0.12)	0.18(0.08)	-0.19(0.08)
30% missing	λ_j	2.90(0.13)	-1.24(0.43)	-0.07(0.45)	-0.98(0.44)
	γ_j	-0.22(0.08)	-0.83(0.17)	0.18(0.10)	-0.19(0.10)
40% missing	λ_j	2.89(0.14)	-1.29(0.46)	-0.07(0.51)	-0.99(0.48)
	γ_j	-0.22(0.08)	-0.81(0.15)	0.19(0.10)	-0.17(0.13)
50% missing	λ_j	2.87(0.16)	-1.34(0.56)	-0.10(0.57)	-1.01(0.56)
	γ_j	-0.22(0.07)	-0.81(0.15)	0.19(0.11)	-0.18(0.15)

Values within parentheses are Monte Carlo standard deviations of 400 estimators

and the truth parameters in these five scenarios.

Figure 16, Figure 17, Figure 18, Figure 19, and Figure 20 with 10 simulations for the above scenarios also show that the the fitted log-innovation variances and generalized autoregressive parameters from unbalanced data using our generalized EM algorithm are close to the true values.

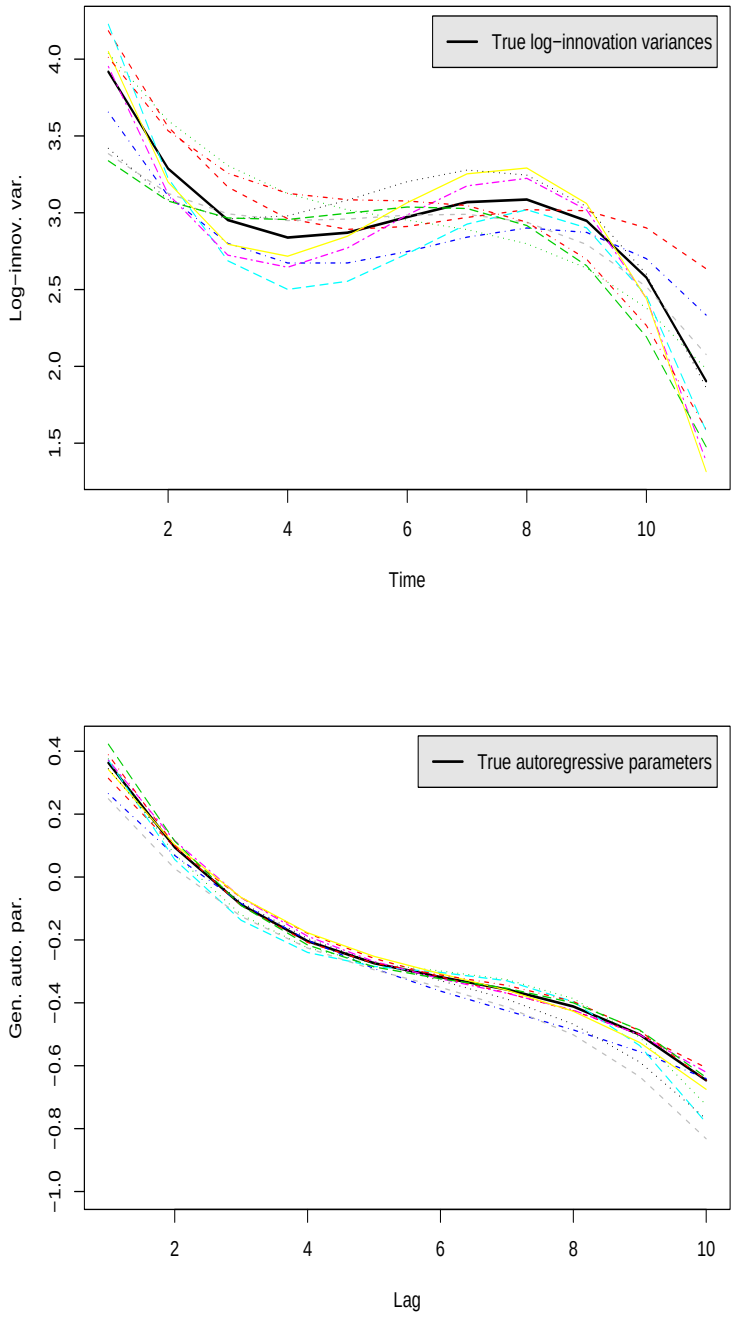

Figure 16: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 10% of the simulation data missing.

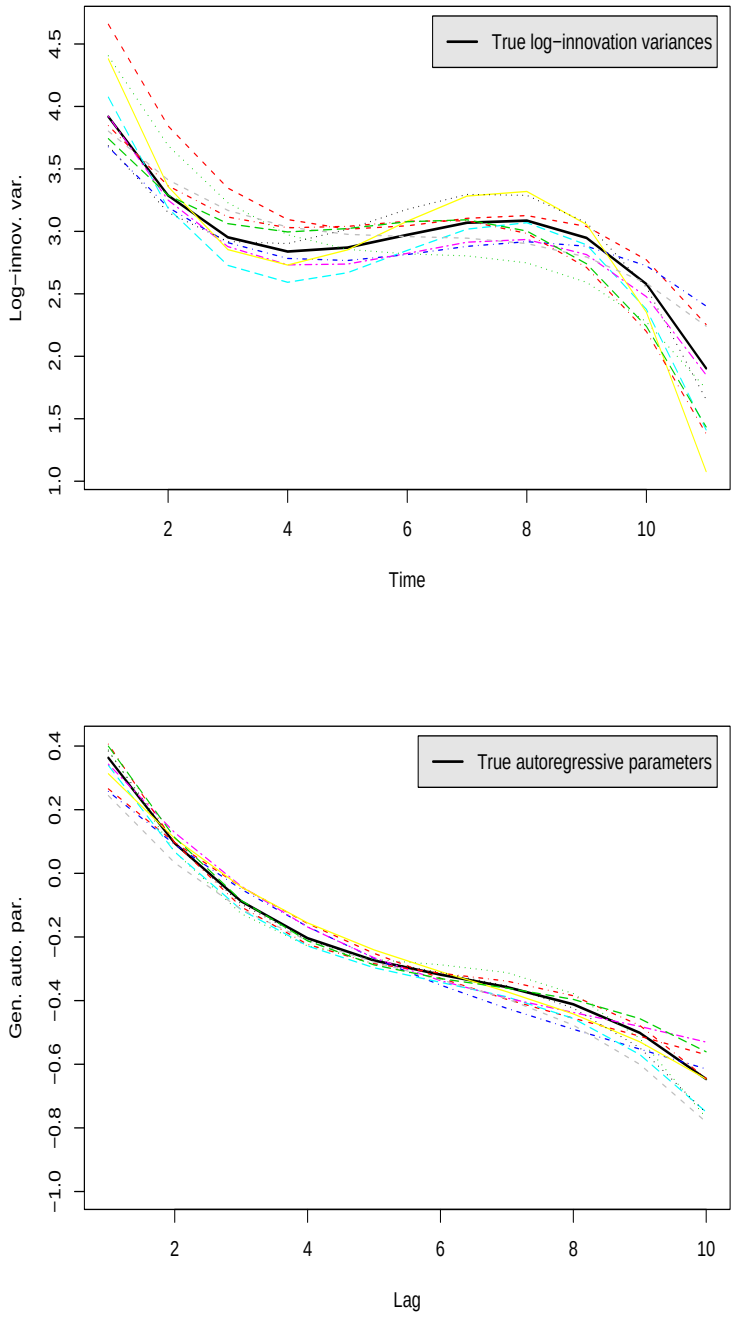

Figure 17: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 20% of the simulation data missing.

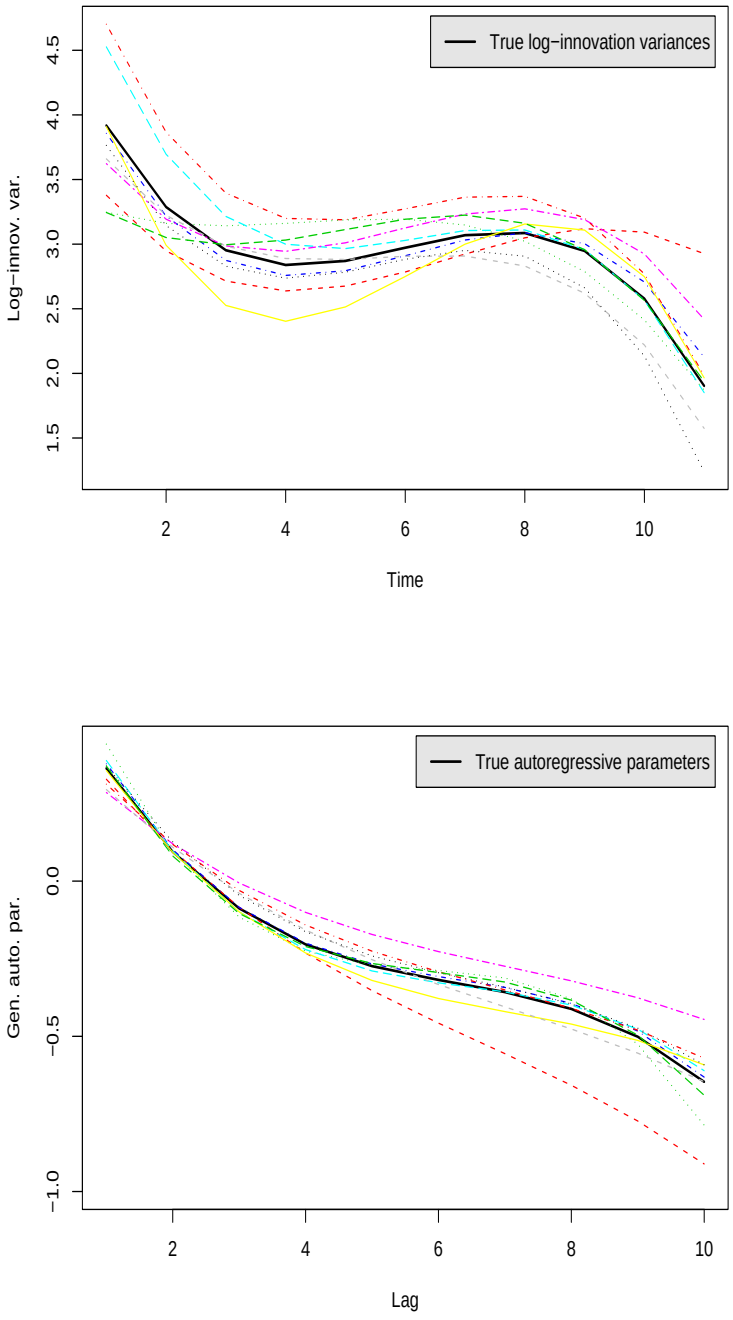

Figure 18: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 30% of the simulation data missing.

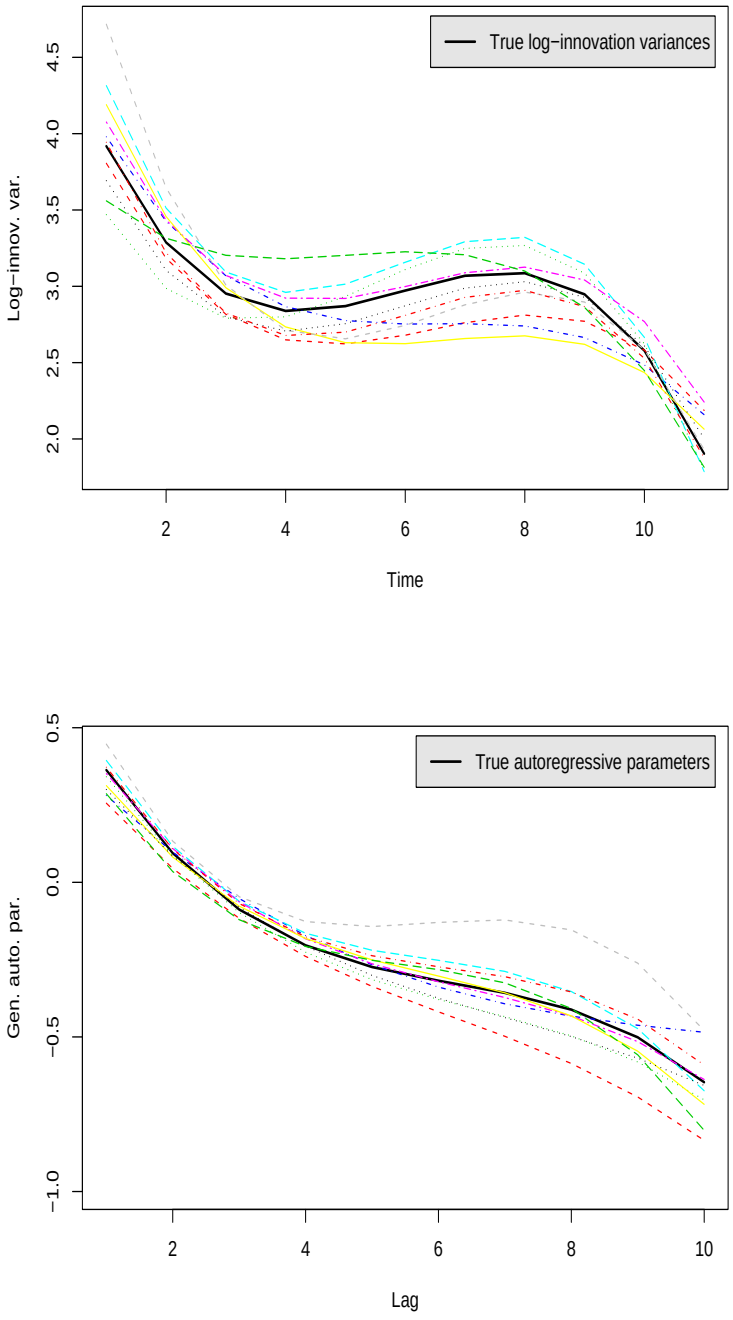

Figure 19: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 40% of the simulation data missing.

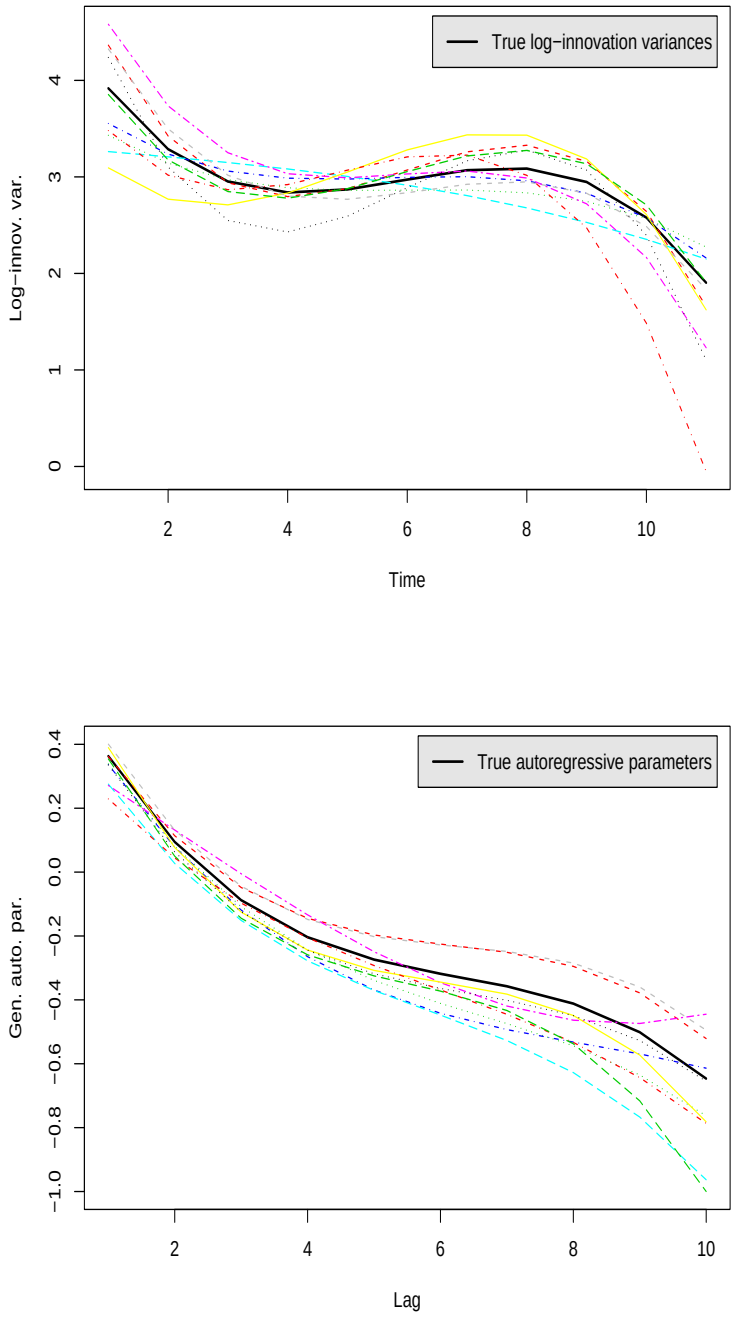

Figure 20: Fitted log-innovation variances and generalized autoregressive parameters for the mixed-effects model with 50% of the simulation data missing.

CHAPTER V

SUMMARY AND FUTURE RESEARCH

5.1 Summary

In this dissertation, we studied reparameterization of covariance matrices with the modified Cholesky decomposition for the fixed-effects and the mixed-effects models for unbalanced longitudinal data. The new parameters have statistical interpretations in terms of innovation variances and autoregressive parameters. The new parameters are not constrained. We can easily model them with covariates in the covariance matrix such that the number of elements to be estimated can be reduced. The modified Cholesky decomposition approach overcomes the difficulties of modeling covariance matrices: positive definite constraint and potential high dimensionality. When observations are not observed at the same set of times, or there are missing observations, we use a Generalized EM algorithm to estimate the ML estimators. We also studied the asymptotic distribution of the ML estimators.

5.2 Future Research

As a future research problem, it would be interesting to extend the proposed methodology to fixed-effects models when the variability over time is influenced by the baseline covariate profile for unbalanced longitudinal data. Furthermore, it would be interesting to extend the proposed methodology to conditionally linear mixed models for unbalanced longitudinal data.

If the covariance matrix is structured, we can not apply this method directly.

Any structure imposed on the covariance matrix will lead to constraints on innovation variances and autoregressive parameters. However, we might handle this problem by choosing appropriate covariates for innovation variances and autoregressive parameters.

REFERENCES

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics* **1**, 135–141.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistics*. New York: Wiley.

Brown, P. J., Le, N. D., and Zidek, J. V. (1994). *Inference for a Covariance Matrix. In Aspects of Uncertainty*. Chichester, U.K.: Wiley.

Chiu, T. Y. M., Leonard, T., and Tsui, K. W. (1996). The matrix-logarithm covariance model. *Journal of the American Statistical Association* **91**, 198–210.

Dempster, A., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.

Diggle, P. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959–971.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24**, 417–441,498–520.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805–820.

Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics* **36**, 296–308.

Laird, N., Lange, N., and Stram, D. (1987). Maximum likelihood computations with repeated measures: Application of the em algorithm. *Journal of the American Statistical Association* **82**, 97–105.

Leonard, T. and Hsu, J. S. J. (1992). Bayesian inference for a covariance matrix. *Annals of Statistics* **20**, 1669–1696.

Munoz, A., Carey, V., Schouten, J. P., Segal, M., and Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* **48**, 733–742.

Newton, H. J. (1988). *Timeslab: A Time Series Analysis Laboratory*. Pacific Grove, CA: Wadsworth & Brooks/Cole.

Oakes, D. (1999). Direct calculation of the information matrix via the em algorithm. *Journal of the Royal Statistical Society, Series B* **61**, 479–482.

Pan, J. X. and MacKenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika* **90**, 239–244.

Pan, J. X. and MacKenzie, G. (2006). Regression models for covariance structures in longitudinal studies. *Statistical Modelling* **6**, 1–15.

Pan, J. X. and MacKenzie, G. (2007). Modelling conditional covariance in the linear mixed model. *Statistical Modelling* **7**, 49–71.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modeling and Inferences Using Likelihood*. Oxford: Oxford University Press.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**, 667–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87**, 425–435.

Pourahmadi, M. and Daniels, M. J. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics* **58**, 225–231.

Szatrowski, T. H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Annals of Statistics* **8**, 802–810.

Vonesh, E. F. and Carter, R. L. (1992). Mixed effects nonlinear regression for unbalanced measures. *Biometrics* **48**, 1–18.

Zimmerman, D. L. and Nunez-Anton, V. (1997). Structured antedependence models for longitudinal data in modelling longitudinal and spatially correlated data: Methods, applications, and future directions. *Springer Lecture Notes in Statistics* **122**, 63–76.

VITA

Min Chen was born in Fuzhou, Fujian, China. She received her B.S. in oceanography from Xiamen University, Fujian, China, June 1996. She received her M.S. in oceanography from Texas A&M University, December 2003. She received her M.S. in statistics from Texas A&M University in August 2005. She received her Ph.D. in statistics from Texas A&M University in May 2008.

Permanent Address:

63, Da Fu Ling, Yi Feng Cun, Xin Dian Xiang, Jin An Qu,

Fuzhou, Fujian, P.R China, 350003